

The quantum Hall curve

J. Nissinen¹ and C.A. Lütken¹

¹*Theory group, Department of Physics, University of Oslo, N-0316 Oslo, Norway*

(Dated: July 20, 2012)

We show how the modular symmetries that have been found to be consistent with most available scaling data from quantum Hall systems^{1,2}, derive from a rigid family of algebraic curves of the elliptic type. The complicated special functions needed to describe scaling data arise in a simple and transparent way from the group theory and geometry of these *quantum Hall curves*. The renormalization-group potential therefore emerges naturally in a geometric context that complements the phenomenology found in our companion paper². We show how the algebraic geometry of elliptic curves is an efficient way to analyze specific scaling data, extract the modular symmetries of the transport coefficients, and use this information to fit the given system into the one-dimensional (real) family of curves that may model all universal properties of quantum Hall systems.

PACS numbers: 73.20.-r

I. INTRODUCTION

Since the discovery of the quantum Hall effect (QHE) three decades ago, a huge amount of experimental data have been collected, but a comprehensive theory that accounts for all universal properties of these systems is still missing.

The situation is somewhat reminiscent of particle physics before the discovery of gauge symmetries. The zoo of “elementary” particles appeared to be organized in simple patterns that turned out to be weight diagrams of some approximate global *unitary* symmetries. The origin of these phenomenological symmetries in the exact underlying gauge theory was only understood much later, but they played an important role in the development of the standard model.

It is still not known how to prove that these symmetries emerge from quantum chromodynamics, but this gap in our understanding of the mathematics of non-abelian gauge theories has not impeded the progress of particle physics. An effective field theory (EFT) known as “the chiral model” can be constructed using the observed low energy degrees of freedom (hadrons) and their approximate global symmetries as input, and this is sufficient to model the low energy dynamics. In order for particle physics to evolve from particle taxonomy to deep underlying principles it was essential to acquire new mathematical tools, including some group theory and geometry.

Similarly, almost all the scaling data in quantum Hall systems appears to be organized in simple patterns that are phase diagrams of some approximate global *modular* symmetries^{1,2}. Their origin must be in the exact underlying gauge theory (quantum electrodynamics in disordered media), but it is not known how to prove that they emerge as effective symmetries at the low energies used in transport experiments. So far efforts to construct an EFT start by postulating some low energy degrees of freedom (see e.g. Refs. 3–6).

It does not seem implausible that modular symme-

tries will aid in the construction of effective field theories that can model the emergent behavior observed in two-dimensional electron liquids, but in order to progress beyond the taxonomy of scaling diagrams it is necessary to acquire new mathematical tools that are not in general use today in condensed matter physics. Our primary purpose here is therefore to explore the mathematical structure underlying the observed modular symmetries, and to show how the data are neatly encoded in the geometry of an algebraic curve that we shall call *the quantum Hall curve*.

These algebraic curves are not without precedent in physics. As is often the case, Nature appears to recycle a good idea, and similar structures appear in integrable models of statistical mechanics. It was Onsager who realized the utility of elliptic functions in this context, which flourished in the Baxter model and its more recent generalizations using algebraic curves (see e.g. Refs. 7,8). They are also the centerpiece of the Seiberg-Witten theory of certain supersymmetric models, where the symmetries are sufficiently rigid to allow the construction of the full low-energy effective action from a certain elliptic curve⁹. In both cases what is required is to have two non-commuting discrete symmetries in parameter space that are combinations of translations and duality transformations. Translation symmetries are ubiquitous in physics, while dualities are not uncommon, being relatives of Kramers-Wannier duality in statistical mechanics, and of electro-magnetic duality in field theory.

In order to establish notation we recall that a quantum field theory is defined by a generating function constructed by summing over all field configurations φ weighted by the piece of the action $S(\varphi; \tau)$ available to the physical system in the experiment under consideration. In other words, the coupling constants $\tau = (\tau_1, \tau_2, \dots)$ span the relevant part \mathcal{M} of the full (infinite dimensional) parameter or *moduli* space of all conceivable field theories. More precisely, in the correct renormalization group (RG) vernacular, \mathcal{M} is the finite dimensional subspace spanned by relevant and marginal operators.

Restricting attention to this finite dimensional moduli space is necessary, in order to do physics, but it is not sufficient. Almost all field configurations cost too much energy, and the effective field theory should retain only those that are accessible at low energy. The conventional (“bottom-up”) approach is to focus first on the effective action, trying to extract some of its properties from the microscopic theory, and subsequently derive some of the properties of \mathcal{M} by calculation. But in quantum Hall experiments it is \mathcal{M} that is probed directly, so our phenomenological (“top-down”) approach is the reverse, as we now summarize in the briefest possible terms.

The symmetries have been extensively discussed in a companion paper², where we analyze available scaling data and show that almost all data fit into a one-parameter family of RG potentials with modular symmetry. The theory of modular groups will therefore not be our main focus here, but rather the underlying geometry to which they are naturally associated. Information on modular groups and their representations has been collected in the Appendix for easy reference.

Since it may not be completely obvious why symmetries, especially intricate modular symmetries, are inextricably entangled with geometry, it may be helpful to recall Felix Klein’s “Erlangen Program”¹⁰. He proposed that abstract symmetries (group theory) organizes geometry, and that projective geometry is the unifying framework. Again the situation is somewhat reminiscent of the inception of gauge theories. Quantum electrodynamics was initially a rather awkward construction based on the obscure idea called “minimal substitution”, which eventually matured into $U(1)$ gauge invariance. It is possible to generalize this symmetry to $SU(n)$ by “brute force”, but this is greatly facilitated by using principal fiber bundles, which is the “Kleinian geometry” where gauge symmetries act naturally. Similarly, the Kleinian geometry to which modular symmetries belong is the complex projective geometry of algebraic curves, and more specifically, the hyperbolic geometry of their moduli spaces.

Consider first the spin polarized QHE. From plateaux and scaling data we first infer that the moduli space is the modular curve $\mathcal{M}_0(2)$ with $\Gamma_0(2)$ symmetry, and that the RG β -function is a holomorphic vector field on this curve¹. Because the RG flow appears to respect both the symmetry and the complex structure of $\mathcal{M}_0(2)$ this is a powerful result: the identification $\mathcal{M} \simeq \mathcal{M}_0(2)$ seems to fix almost all global universal data, and we are led to consider the family $\mathcal{E}_0(2)$ of elliptic curves whose moduli space is $\mathcal{M}_0(2)$.

So far the geometry of these elliptic curves, which are “rigidified” (also called “framed”, “enhanced” or “decorated” in the mathematical literature) by certain torsion data, have received little attention in physics, apart from their relation to spin structures on the torus. One reason is that so much of the physically relevant data about universality classes is encoded directly in the topology and geometry of the moduli space. In the polarized case,

$\mathcal{M}_0(2)$ is so rigid that it predicts the exact location of all quantum critical points, as well as the precise shape of all RG flow lines. Remarkably, this is in agreement with a host of experiments. One of the coldest experiments to date¹¹, where the effective field theory is expected to be most accurate, has verified the modular prediction from 1992 at the *per mille* level¹.

There is no obvious physical reason why the phenomenological approach should be restricted to the polarized system, provided we are willing to consider different symmetries, which usually will be smaller than $\Gamma_0(2)$ (but never larger). The fully polarized case is continuously connected to the unpolarized system by tuning one or more suitable control parameters, at least in principle. This raises the question how these “symmetry transitions”, which appear to be discontinuous, can come about. We address this question by studying a family of algebraic curves, which gives rise to a surprisingly simple family of RG potentials Φ_a that interpolates between the points of enhanced symmetry.

Another compelling reason to study these curves is to understand the origin and properties of holomorphic vector fields on \mathcal{M} , which include the β -functions whose integral curves are the RG flow lines in these models. In maximally symmetric cases they are essentially unique, and less symmetric cases at level 2 are only slightly less constrained. The additional freedom is sufficient, but just barely so, to account for unpolarized data. This is one example of a remarkable confluence of quantum Hall physics and modular mathematics, which is best appreciated in a geometric setting.

The absence of a derivation of these symmetries from a microscopic model of charge transport makes it difficult to compare our phenomenological analysis with more conventional models. This applies to all three regimes of interest: the infrared (IR) plateaux domain where Laughlin-like wave functions and Chern-Simons type topological theories capture the relevant physics; the ultraviolet (UV) perturbative domain where non-linear sigma models³ describe the relevant modes (for a current review see Ref. 4); and the scaling domain where a conformal field theory is expected to appear, albeit a rather subtle one. The UV/IR limits are at the boundary of \mathcal{M} , infinitely far away from the scaling region in any natural metric on moduli space. This is, perhaps, where one would expect our effective field theory to break down, and be replaced by other models better adapted to the UV/IR limits. Indeed, using the algebraic approach advocated here, we will show that quantum Hall curves degenerate to singular geometries in these limits. Our approach has been tailored for the quantum critical regions controlled by quantum critical points deep inside moduli space. A deep understanding of the geometry of the quantum Hall curves at these points may aid in the construction of the conformal quantum critical model.

The next section is a primer on elliptic curves that provides the theoretical framework and mathematical tools

required for the experimental analysis, presented in a companion paper². One of the main goals in this section is to clarify the algebraic and geometric origin of the special (elliptic and modular) functions that we use extensively. We also explain the geometric origin of families of curves that have modular symmetries smaller than the full modular group $\text{PSL}(2, \mathbb{Z})$, and we discuss all congruence sub-groups at level two. This appears to be sufficient for the analysis of unpolarized quantum Hall data².

In Sec. III we turn our attention to a more detailed discussion of the modular curve \mathcal{M} . Modular functions that live on the modular curve \mathcal{M} arise from elliptic functions that live on the elliptic curve \mathcal{E} , encountered in the previous section.

This equips us with the technology to explain in Sect. IV how scaling experiments inform us of local and global properties of the renormalization group, and how to reconcile this (semi-)group with modular symmetries. The resulting β -functions are rather complex holomorphic modular vector fields with the remarkable property that they generate gradient flows. This is both physically interesting, since RG potentials are expected to exist on general grounds, and also a great mathematical simplification, since we need only consider scalar fields on \mathcal{M} . Combined with some reasonable physical assumptions about RG flows in general, this finally leads us to propose a simple one parameter family of curves with RG potentials Φ_a invariant under level two modular symmetries. Our data analysis² indicates that this potential is consistent with almost all available scaling data, for $a \in \mathbb{R}$.

II. ELLIPTIC CURVES

The modular symmetry observed in the QHE is intimately related to a special type of Riemann surface, which belongs to a family of so-called *elliptic curves*. In the jargon of nineteenth century geometry, this curve is a “1-pointed genus-1 curve with 2-torsion”. The unfortunate use of both “curve” and “surface” for the same object derives from the convention in algebraic geometry that curves by default are defined over \mathbb{C} , while Riemann surfaces are over \mathbb{R}^2 . So a “curve” has one complex dimension but two real dimensions. The genus g is the number of holes in this real 2-dimensional surface. The rôle of one (origin) or more (torsion) privileged points is explained below. In order to understand the properties of these models, in particular the existence and uniqueness of the holomorphic modular vector fields used extensively in our analysis of the scaling data, we need a primer on this geometric structure, which we now supply.

A. Topology of tori

The flat 2-dimensional torus T^2 is by definition the surface obtained from \mathbb{R}^2 by the identifications

$$(x, y) \simeq (x + 1, y) \simeq (x, y + 1). \quad (1)$$

In other words, topologically a torus is just the product of two circles, $T^2 \simeq S^1 \times S^1$.

Any torus can be mapped to any other by a continuous change of real coordinates, so unless we endow them with some additional structure they are rather featureless. Demanding smoothness does not refine this classification: the genus g is the only invariant under homeomorphisms or diffeomorphisms. In order to model the QHE quite a lot of “decoration” with extra data is needed: the additional data are called *complex structure* and *torsion*.

We first explain what a complex structure is. Two tori are only equivalent, i.e., diffeomorphic, as complex manifolds if there is a holomorphic (complex analytic) coordinate transformation between them, and this is often not possible. Two tori that are distinct in this sense are said to have different complex structures. One way to parametrize this structure is by introducing the complex coordinate $z := x + \tau y$ ($\text{Im } \tau \neq 0$) on the torus. The periodic identifications dictated by the topology are now

$$z \simeq z + 1 \simeq z + \tau. \quad (2)$$

In other words, we can represent the torus as the complex plane “rolled up” in both directions, i.e., $T_\tau^2 = \mathbb{C}/\Lambda_\tau$, where $\Lambda_\tau = \mathbb{Z} + \mathbb{Z}\tau$ is the skewed lattice in \mathbb{C} shown in Fig. 1 (a) with basis vectors $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, \tau)$. By choosing an ordering of the basis we can limit $\text{Im } \tau > 0$. Notice that we have already used the “small” complex diffeomorphisms of the torus: a translation to fix the origin of the lattice, and a scaling to normalize the first basis vector. It is obvious that the geometry and topology of a genus-1 curve is that of a torus, compare Fig. 1 (b). Once an origin of the lattice has been singled out, the torus is called *an elliptic curve*.

It is clear that τ labels all possible complex structures, but in a highly redundant way. We can restrict τ without loss of generality to the upper half of the complex plane,

$$\tau \in \mathbb{H} := \{\tau \in \mathbb{C} | \text{Im } \tau > 0\},$$

but this still leaves an infinite number of “copies” T_τ^2 , of each complex torus T_τ^2 . In fact, τ' represents the same lattice, i.e., the same complex structure, as τ iff $\tau' = \gamma(\tau)$, where γ is a Möbius transformation taken from the modular group $\Gamma(1) = \text{PSL}(2, \mathbb{Z})$. These are the “large” diffeomorphisms of the torus. How this comes about in terms of equations is detailed below, when considering elliptic curves. There is one last discrete diffeomorphism of the torus with the complex structure τ in the upper half plane \mathbb{H} , which corresponds to changing the orientation of Λ_τ , acting as $\tau \mapsto -\bar{\tau}$ and not a modular transformation. Instead, it is an outer automorphism

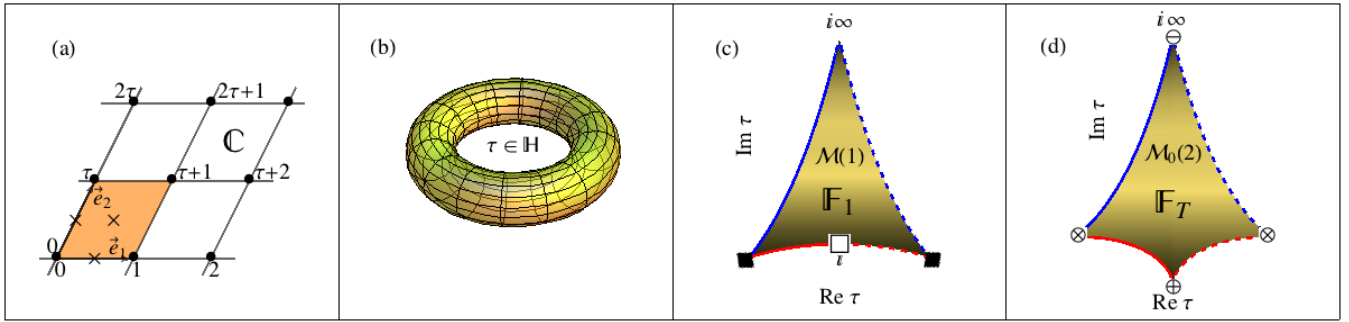


Figure 1: (Color online) An *elliptic curve* is both (a) a lattice Λ_τ (crosses are the 2-torsion points mod Λ_τ), and (b) a genus $g = 1$ Riemann surface. (c) In-equivalent elliptic curves are uniquely parametrized by the points of the hyperbolic triangle \mathbb{F}_1 , which is the interior of the *modular curve* $\mathcal{M}(1)$ of genus $g = 0$. (d) The polarized quantum Hall curve $\mathcal{E}_0(2)$ has symmetry $\Gamma_0(2)$ with fundamental domain \mathbb{F}_T , and is parametrized by the modular curve $\mathcal{M}_0(2)$.

of $\Gamma(1)$, and arises in the QHE as the particle-hole conjugation. Since this is simply a reflection in the imaginary axis of \mathbb{H} , its consequences are rather simple and we will focus on the modular transformations.

The modular group partitions \mathbb{H} into equivalence classes, and a region of \mathbb{H} that represents each possible complex structure precisely once is called a fundamental domain \mathbb{F}_1 of $\Gamma(1)$. The simplest choice of this domain is to take τ above the unit circle in the strip $|\operatorname{Re} \tau| \leq 1/2$,

$$\mathbb{F}_1(\tau) = \{\tau \in \mathbb{H} \mid -1/2 \leq \operatorname{Re} \tau < 1/2, |\tau| \geq 1\}.$$

Identifying all points at infinity, we can think of the fundamental domain as the hyperbolic triangle shown in Fig. 1 (c). Identifying points on the boundary of \mathbb{F}_1 that are connected by modular transformations gives the moduli space $\mathcal{M}(1)$. Topologically it is also a curve, and in fact an algebraic curve of genus-0, i.e., the Riemann sphere, which is called *the modular curve* for the full modular group.

Adding more structure breaks the symmetry to a subgroup $\Gamma \subset \Gamma(1)$, splitting up the equivalence classes further so that a larger fundamental domain is needed in order to label them all. For example, in the spin-polarized quantum Hall case the modular symmetry is $\Gamma_0(2)$, which has the fundamental domain

$$\mathbb{F}_T(\tau) = \{\tau \in \mathbb{H} \mid -1/2 \leq \operatorname{Re} \tau < 1/2, |\tau \pm 1/2| \geq 1/2\}.$$

Identifying all points at infinity, we can think of the fundamental domain as the hyperbolic polygon shown in Fig. 1 (d). Identifying points on the boundary of \mathbb{F}_T that are connected by modular transformations taken from $\Gamma_0(2)$ gives the level 2 moduli space $\mathcal{M}_0(2)$. Topologically it is also a curve, and again an algebraic curve of genus-0, which is called the modular curve for the congruence subgroup $\Gamma_0(2) \subset \Gamma(1)$, sometimes abbreviated to “submodular curve”. We shall soon meet other parameterizations of the complex structure of elliptic curves.

B. Algebraic curves

In order to explain the additional structure (torsion) needed to model the QHE, we turn to the algebraic representation of elliptic curves. It is far from obvious that an *algebraic curve*, i.e., a dimension one subset of a complex projective space, has anything to do with two-dimensional tori (elliptic curves).

The simplest algebraic presentation of a torus is as a “planar cubic”. The complex projective “plane” \mathbb{CP}^2 has four real dimensions, which is reduced to two real dimensions by any sufficiently general (“complete”) algebraic (polynomial) constraint, since the polynomial is also defined over \mathbb{C} . The hypersurface defined by the vanishing locus of the polynomial is therefore a Riemann surface, i.e., a complex curve. If the polynomial has degree three then it is an elliptic curve.

There are infinitely many other ways to present elliptic curves, by algebraic constraints in ambient projective spaces of any dimension, but so-called “complete intersections” are especially simple and therefore especially useful. There are only two: planar cubics and the intersection of two quadrics in \mathbb{CP}^3 . The black skewed squares in Fig. 1 (c) are pictograms of tori with these shapes, which are simply related to planar cubics. The white square icon at $i = \sqrt{-1}$ in Fig. 1 (c) is a pictogram of a torus with this shape, which is simply related to the intersection of two quadrics.

Independent of any specific presentation, an elliptic curve has two remarkable properties. First, it is always isomorphic to a torus, and any (marked) torus is isomorphic to an elliptic curve. This means that there is always a (highly transcendental) correspondence between the coefficients of the defining polynomials and the complex number τ labeling tori.

Second, unlike all other curves ($g \neq 1$), a genus-1 curve can be endowed with a group structure. This is the source of its remarkable arithmetic properties that even after two centuries of intense scrutiny remains an active field

of research. More precisely, a genus-1 curve is always an *abelian group* provided one point (in the following construction, the point O at infinity) is chosen as the zero element for the “chord-tangent” group law, which to any two points P and Q on the curve gives a third point $P+Q$. The freedom in choosing the identity of this composition rule corresponds to the freedom to translate the toroidal lattice Λ_τ to choose any point to serve as the origin, since the abelian group law is equivalent to simply addition in \mathbb{C}/Λ_τ . By definition an elliptic curve comes equipped with this point, but a topological torus without a complex structure does not.

C. Planar cubics

By a suitable change of coordinates we can always bring a cubic polynomial in \mathbb{CP}^2 to the form $y^2 = \mathcal{E}(x)$, where $\mathcal{E}(x)$ is a cubic polynomial in x , and $x = X/Z$ and $y = Y/Z$ are inhomogeneous coordinates obtained from the homogeneous coordinates X, Y and Z of \mathbb{CP}^2 on a chart with a constant $Z \neq 0$. In order to see this we first observe that a general cubic constraint with a marked point O can be written¹³

$$ZY^2 + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

after a suitable rescaling. This curve has the special point O at infinity given by $X = Z = 0$ and $Y \neq 0$. Since $Z \neq 0$ elsewhere on the curve, the inhomogeneous form of this equation is

$$\mathcal{E} : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \quad (3)$$

To what extent is this equation unique, i.e., how does the elliptic curve \mathcal{E} (with O fixed at infinity) depend on the constants a_1, a_2, a_3, a_4 and a_6 ? An *admissible* change of variables, i.e., a coordinate transformation which preserves the above form and retains the special point O at infinity, is given by

$$x = u^2x' + r, \quad y = u^3y' + su^2x' + t, \quad u \neq 0. \quad (4)$$

The composition of two admissible changes of variables is again admissible. Since we work over \mathbb{C} , we can use the admissible coordinate transformation $y \mapsto (y - a_1x - a_3)/2$ to complete the square, bringing the equation to the much simpler form

$$y^2 = 4x^3 + b_2x^2 + b_4x + b_6.$$

Another admissible coordinate transformation given by $(x, y) \mapsto (x - b_2/12, y)$ eliminates the quadratic term, leaving us with a deceptively simple looking two-parameter family of cubics:

$$\mathcal{E}_{A,B} : y^2 = 4x^3 + Ax + B.$$

The only change of variables that preserves this form is

$$(x, y) \mapsto (u^2x, u^3y), \quad (5)$$

which suggests that the invariant combination

$$j \propto \frac{A^3}{A^3 + 27B^2}$$

may be useful. Whatever the transcendental connection is between the complex structure τ of the torus and the algebraic complex coefficients A and B , this so-called “ j -invariant” is well defined as long as $A^3 + 27B^2 \neq 0$.

Retracing the steps back to the general cubic in Eq. (3), it is a long but straightforward calculation to show that the j -invariant is indeed preserved, i.e., that $j' = j$ under any admissible change of variables given by Eq. (4). Moreover, all changes of variables between the different forms of the cubics are admissible, so one can work with any of the above forms for a given curve. These presentations of the cubic are collectively known as “the Weierstrass form”.

As already mentioned, an elliptic curve \mathcal{E} is also defined by a lattice $\Lambda(\omega_1, \omega_2) = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ in \mathbb{C} :

$$\begin{aligned} \mathcal{E} &\cong \mathbb{C}/\Lambda(\omega_1, \omega_2) \\ &= \{z \simeq z + m\omega_1 + n\omega_2 \mid z \in \mathbb{C}; m, n \in \mathbb{Z}\}, \end{aligned}$$

where $\omega_1/\omega_2 \notin \mathbb{R}$, and we can take $\text{Im } \omega_1/\omega_2 > 0$ without loss of generality. But two different lattices can give rise to an equivalent Weierstrass form under admissible change of variables.

First, two bases (ω_1, ω_2) and (ω'_1, ω'_2) span the same lattice iff

$$\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix}, \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1).$$

Two lattices Λ and Λ' will be considered equivalent if $\Lambda' = \alpha\Lambda$ with $\alpha \in \mathbb{C}^\times$ (we write this as $\Lambda \sim \Lambda'$), since it will transpire that the corresponding elliptic curves are isomorphic, i.e., have the same Weierstrass form. We have seen how to use this freedom to determine equivalent complex structures τ of the torus, where they arise as complex diffeomorphisms. That is, we can always rescale with ω_2 , so that

$$\Lambda(\omega_1, \omega_2) \sim \Lambda(1, \tau = \omega_1/\omega_2) = \Lambda_\tau = \mathbb{Z} + \mathbb{Z}\tau,$$

where $\text{Im } \tau > 0$ without loss of generality. From this it follows that

$$\Lambda_{\tau_1} \sim \Lambda_{\tau_2} \iff \tau_2 = \gamma(\tau_1) = \frac{a\tau_1 + b}{c\tau_1 + d}$$

with $\gamma \in \Gamma(1)$ and $\tau_1, \tau_2 \in \mathbb{H}$. To see this, notice that $\Lambda_{\tau_1} = \alpha\Lambda_{\tau_2} = \mathbb{Z}\alpha + \mathbb{Z}\alpha\tau_2$, which means that $\alpha \cdot 1 = c\tau_1 + d$ and $\alpha\tau_2 = a\tau_1 + b$, so $\tau_2 = \gamma(\tau_1)$ and $\alpha = c\tau_2 + d$. The converse follows by choosing $\alpha = c\tau_2 + d$, and using that the resulting bases are then related by a $\Gamma(1)$ transformation.

An explicit form for the transcendental transformation between the complex structure parameter τ (the lattice

Λ_τ) and the elliptic curve \mathcal{E} in Weierstrass form, is given by the Weierstrass \wp -function

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda/\{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right), \quad (6)$$

which is invariant to lattice transformations, $\wp_\Lambda(z+\Lambda) = \wp_\Lambda(z)$. The \wp -function satisfies the differential equation

$$(\wp'_\Lambda)^2 = 4\wp_\Lambda^3 - g_2(\Lambda)\wp - g_3(\Lambda). \quad (7)$$

With $x = \wp$ and $y = \wp'$ we therefore have an explicit embedding of the torus \mathbb{C}/Λ_τ into the complex projective plane, provided we can always find Λ so that $g_2(\Lambda) = -A$ and $g_3(\Lambda) = -B$.

Two curves will have identical j -invariants if $\Lambda' = \alpha\Lambda$, since $\wp_{\alpha\Lambda}(\alpha z) = \alpha^{-2}\wp_\Lambda(z)$ so $(x, y) \mapsto (\alpha^{-2}x, \alpha^{-3}y)$, which is an admissible change of variables with $j = j'$. We have already seen that this means that $\Lambda_{\tau'} = (c\tau + d)\Lambda_\tau$, so we immediately have the properties

$$g_2\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^4 g_2(\tau) \quad (8)$$

$$g_3\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^6 g_3(\tau) \quad (9)$$

$$j\left(\frac{a\tau+b}{c\tau+d}\right) = j(\tau). \quad (10)$$

The coefficients $g_2(\tau) := g_2(\Lambda_\tau)$ and $g_3(\tau) := g_3(\Lambda_\tau)$ are modular forms of weight 4 and 6, respectively, and they generate the ring of all modular forms for $\Gamma(1)$ to be discussed below. Using the definition of $j(\tau)$ and the modular properties of j , g_2 and g_3 , one can complete the equivalence of elliptic curves \mathcal{E}_τ and lattices Λ_τ by showing that there always exist a Λ_τ such that $g_2(\tau) = -A$ and $g_3(\tau) = -B$.¹³

The important point is that we see how the elliptic and modular functions parametrize the elliptic curve, i.e., the torus with complex structure and origin, and how their transformation properties respect the invariances of these objects.

In summary, the following objects are equivalent:

- i) Tori with complex structure τ .
- ii) Lattices $\{\Lambda_\tau = \mathbb{Z} + \mathbb{Z}\tau \mid \tau \in \mathbb{H}/\Gamma(1)\}$ in \mathbb{C} .
- iii) Cubic equations $y^2 = 4x^2 + Ax + B$ in \mathbb{CP}^2 with $j(\tau) \propto A^3/(A^3 + 27B^2)$.

The modular curve $\mathbb{H}/\Gamma(1)$ is a moduli space of elliptic curves, i.e., it is a space that parametrizes equivalence classes of curves that are isomorphic under modular transformation of the lattice and admissible coordinate transformations that preserve the origin. The modular invariant function $j(\tau)$ takes different values for different (equivalence classes of) elliptic curves.

In the following we shall also discuss subgroups $\Gamma(2) \subseteq \Gamma \subset \Gamma(1)$ of the modular group. The (sub-)modular

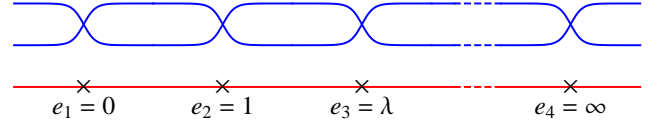


Figure 2: (Color online) Branched covering of the elliptic curve in \mathbb{CP}^2 over \mathbb{CP}^1 , with branch points e_1 , e_2 , e_3 and e_4 (adapted from a diagram on p.186 in Ref. 12).

curves \mathbb{H}/Γ are moduli spaces of elliptic curves with additional structure, and these will have their own special functions. This structure is intimately related to the inverse of the \wp -map discussed above, taking a torus to an elliptic curve. It can be shown¹² that elliptic curves are mapped to tori by the complex valued hypergeometric function $F(1/2, 1/2, 1; \lambda)$. This is in general a two-to-one map, as illustrated in Fig. 2, except at the branch points e_i , which have special group theoretical properties called “2-torsion”.

D. Torsion points

The simplest way to understand the abelian group law, and the special points on \mathcal{E}_τ related to this, is to consider addition modulo Λ_τ on \mathbb{C} . The embedding $\mathcal{E}_\tau \hookrightarrow \mathbb{CP}^2$ defined by $(x, y, 1) = (\wp_{\Lambda_\tau}(z), \wp'_{\Lambda_\tau}(z), 1)$ is well defined since \wp_{Λ_τ} is Λ_τ -periodic. In particular, the elliptic curve \mathcal{E}_τ has N -torsion points that satisfy $Nz = 0 \pmod{\Lambda_\tau}$. These points correspond to rational points $\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}\tau$ on \mathcal{E}_τ , and they are the elements of the torsion group $\mathcal{E}_\tau[N]$.

We will only need to consider 2-torsion points, i.e., points that satisfy $2z = 0 \pmod{\Lambda_\tau}$ or $z = -z \pmod{\Lambda_\tau}$, which have a particularly simple description. Clearly the non-trivial 2-torsion points for \mathbb{C}/Λ_τ are

$$z_1 = \frac{1}{2}, \quad z_2 = \frac{\tau}{2}, \quad z_3 = \frac{1+\tau}{2},$$

which are the crossed points in Fig. 1 (a). Since $\wp(-z) = \wp(z)$ and $\wp'(-z) = -\wp'(z)$, the 2-torsion points $P = -P$ in \mathbb{CP}^2 satisfy

$$P = (x, y, 1) = -P = (x, -y, 1), \quad P \in \mathcal{E}_\tau.$$

The 2-torsion points are therefore given by the roots of the planar cubic \mathcal{E}_τ ,

$$y^2 = 4x^3 + Ax + B = 4(x - e_1)(x - e_2)(x - e_3), \quad (11)$$

and $(\wp(z_i), \wp'(z_i)) = (e_i, 0)$, with $e_1 + e_2 + e_3 = 0$ since the quadratic term vanishes. The roots are distinct iff

$$\Delta = -(A^3 + 27B^2) = 16(e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2 \neq 0.$$

Consider now, for the sake of generality, how the isomorphisms $\mathcal{E}_\tau \rightarrow \mathcal{E}_{\gamma(\tau)}$ induced by transformations $\gamma \in \Gamma$

act on N -torsion points. We have seen that for any $\gamma \in \Gamma(1)$, the isomorphism acts as

$$f_\gamma : \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau) \mapsto \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\gamma \cdot \tau), \quad z \mapsto \frac{z}{c\tau + d}.$$

We discuss the level N groups $\Gamma_0(N)$, $\Gamma^0(N)$, $\Gamma_\theta(N)$, and $\Gamma(N)$ in turn, following Ref. 13.

Consider first how $\gamma \in \Gamma_0(N)$, which has $c = 0 \pmod{N}$ and $a = d = 1 \pmod{N}$, acts on the cyclic subgroup of order N given by

$$C_N = \left\{ \frac{1}{N}, \dots, \frac{N-1}{N} \right\} \subset \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau),$$

with $\tau \in \mathbb{H}/\Gamma_0(N)$. We see that $1/N \mapsto 1/N(c\tau + d)$, and it follows that

$$\begin{aligned} \frac{k}{N} - \frac{k}{N(c\tau + d)} &= \frac{k}{N} \frac{c\tau + d - 1}{c\tau + d} \\ \in f_\gamma(\mathbb{Z} + \mathbb{Z}\tau) &= \mathbb{Z} + \mathbb{Z}\gamma \cdot \tau = 0, \end{aligned}$$

for any $k/N \in C_N$ and $k = 1, \dots, N-1$. Thus the points k/N and $k/N(c\tau + d)$ are equivalent mod $\Lambda_{\gamma \cdot \tau}$ and map to each other in $\mathbb{C}/(\mathbb{Z} + \mathbb{Z}\gamma \cdot \tau)$.

We have seen that $\mathbb{H}/\Gamma(1)$ parametrizes equivalence classes of elliptic curves that are isomorphic under an admissible coordinate transformation that need only preserve one point, namely the origin which “rigidifies” a torus to an elliptic curve. Similarly, $\mathbb{H}/\Gamma_0(N)$ is a moduli space for the equivalence classes of (\mathcal{E}_τ, P_N) , where P_N is a point of order N . Any point in C_N is fixed and P_N generates this group. In other words, $\mathbb{H}/\Gamma_0(N)$ parametrizes elliptic curves up to coordinate and modular transformations that preserve both the origin and this additional torsion point of order N , further “rigidifying” what we mean by an elliptic curve.

The group $\Gamma^0(N)$, for which $b = 0 \pmod{N}$ and $a = d = 1 \pmod{N}$, carries the points of the cyclic subgroup $C_N^\tau = \{\tau/N, \dots, \tau(N-1)/N\}$ to themselves. The moduli space $\mathbb{H}/\Gamma^0(N)$ labels elliptic curves $(\mathcal{E}_\tau, P_N^\tau)$, and $\mathbb{H}/\Gamma^0(N) \neq \mathbb{H}/\Gamma_0(N)$ because the two moduli spaces are “tagged” with different torsion points, $P_N^\tau \neq P_N$.

The group $\Gamma_\theta(N)$ has $b = c = 0 \pmod{N}$ and $a = d = 1 \pmod{N}$ or *vice versa*. It therefore leaves the points $(1 + \tau)k/N$ fixed for all N and $k = 1, \dots, N-1$, which is another cyclic subgroup of $\mathcal{E}_\tau[N]$, see Fig. 1 (a).

$\Gamma(N)$ has $b = c = 0 \pmod{N}$ and $a = d = 1 \pmod{N}$, from which it follows that

$$\begin{aligned} \frac{1}{N} - \frac{1}{N(c\tau + d)} &= \frac{c\tau + d - 1}{N(c\tau + d)} \\ \in f_\gamma(\mathbb{Z} + \mathbb{Z}\tau) &= \mathbb{Z} + \mathbb{Z}\gamma \cdot \tau = 0 \end{aligned}$$

and

$$\begin{aligned} \frac{1}{N} \frac{a\tau + b}{c\tau + d} - \frac{\tau}{N(c\tau + d)} &= \frac{(a-1)\tau + b}{N(c\tau + d)} \\ \in f_\gamma(\mathbb{Z} + \mathbb{Z}\tau) &= \mathbb{Z} + \mathbb{Z}\gamma \cdot \tau = 0. \end{aligned}$$

A point in $\mathbb{H}/\Gamma(N)$ is an elliptic curve \mathcal{E}_τ with a basis $\{1/N, \tau/N\}$ for the N -torsion subgroup $\mathcal{E}_\tau[N]$.

For $N = 2$ these cyclic subgroups have only one point that remains fixed under the morphisms $\mathcal{E}_\tau \rightarrow \mathcal{E}_{\gamma \cdot \tau}$ from the index 2 subgroups. The other points in $\mathcal{E}_\tau[2]$ are interchanged, because torsion points map to torsion points. Since the common subgroup $\Gamma(2)$ preserves all the $\mathcal{E}_\tau[2]$ -points, the information about these points is the extra structure on the elliptic curve that distinguishes the moduli spaces \mathbb{H}/Γ . We have seen that the 2-torsion points correspond to the roots e_1 , e_2 and e_3 of the cubic equation for \mathcal{E}_τ .

E. Legendre curves

Subjecting the Weierstrass form given in Eq. (11) to the change of variables $y \rightarrow y/2$, $\Delta \neq 0$ and

$$x = (e_2 - e_1)x' + e_1, \quad y = (e_2 - e_1)^{3/2}y',$$

we get

$$y^2 = x(x-1) \left(x - \frac{e_3 - e_1}{e_2 - e_1} \right).$$

Any elliptic curve can therefore be given in the so-called *Legendre form*:

$$\mathcal{E}_\lambda : y^2 = x(x-1)(x-\lambda), \quad (12)$$

where the $\Gamma(2)$ -invariant function

$$\lambda = \frac{e_3 - e_1}{e_2 - e_1} \neq 0, 1, \infty$$

parametrizes the isomorphism classes of elliptic curves with fixed basis of the 2-torsion group. We need to understand in more detail the relation between the curves given in eqs. (7) and (12).

As we have already discussed, a level 2 curve carries information about the 2-torsion group $\mathcal{E}_\tau[2]$. Modular transformations that are in the coset $\Gamma(1)/\Gamma(2) \cong S_3$ permute these torsion points, or equivalently the roots of the cubic polynomial of the curve. It is instructive to make this explicit.

Assume that two level 2 curves \mathcal{E}_λ and $\mathcal{E}_{\lambda'}$ are isomorphic as elliptic curves, i.e., that $j(\mathcal{E}_\lambda) = j(\mathcal{E}_{\lambda'})$. Then we know that their Legendre equations are related by a change of variables of the form

$$x = u^2x' + r, \quad y = u^3y',$$

which gives

$$\begin{aligned} y'^2 &= x'(x' - 1)(x' - \lambda') \\ &= \left(x' + \frac{r}{u^2} \right) \left(x' - \frac{1-r}{u^2} \right) \left(x' - \frac{\lambda-r}{u^2} \right). \end{aligned}$$

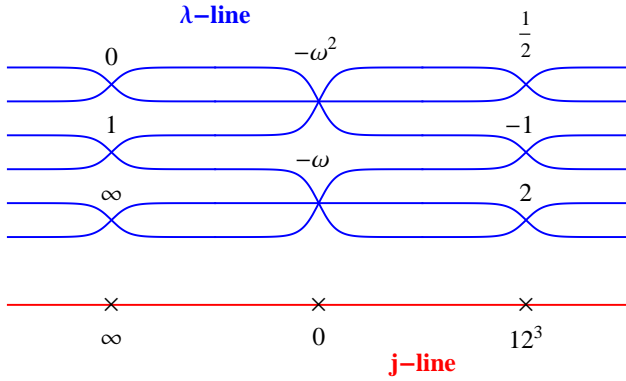


Figure 3: (Color online) The λ -line is a six-to-one branched covering over the j -line. At branch points ($\omega = e^{2\pi i/3}$ is a cube root of unity) the curve has extra automorphisms, and the symmetry is enhanced to subgroups between $\Gamma(2)$ and $\Gamma(1)$ (adapted from a diagram on p.88 in Ref. 12).

There are six possible solutions for the roots $e_i(r, u^2, \lambda)$ ($i = 1, 2, 3$) in terms of permutations of $0, 1, \lambda'$, and we get

$$\lambda' \in \left\{ \lambda, 1 - \lambda, \frac{1}{\lambda}, \frac{1}{1 - \lambda}, \frac{\lambda}{\lambda - 1}, \frac{\lambda - 1}{\lambda} \right\}. \quad (13)$$

We see that the parametrization in terms of λ is six-to-one in terms of j , as illustrated in Fig. 3, corresponding to the index of $\Gamma(2)$ in $\Gamma(1)$. Indeed, if one writes j in terms of e_1, e_2 and e_3 , the expression is invariant under permutations of the roots. Modular transformations in $\Gamma(1)/\Gamma(2) \cong S_3$ permute the roots of the equation, and $\lambda = (e_3 - e_1)/(e_2 - e_1)$ transforms like the cross ratio of the points e_1, e_2, e_3 and $e_4 = i\infty$ on the coordinate $x \cong \mathbb{CP}^1$. Eq. 13 is therefore a representation of S_3 in terms of the modular transformations in the coset $\Gamma(1)/\Gamma(2)$.

These properties allow us to write j in terms of λ . The change of variables:

$$(x, y) \rightarrow \left(x + \frac{\lambda + 1}{3}, \frac{y}{2} \right)$$

takes the Legendre curve to the Weierstrass form

$$y^2 = 4x^3 - \frac{4}{3}(\lambda^2 - \lambda + 1)x - \frac{4}{27}(\lambda + 1)(\lambda - 2)(2\lambda - 1),$$

and it follows that

$$j(\lambda) = 2^8 \frac{(\lambda^3 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}.$$

This non-singular as long as $\lambda \neq 0, 1, \infty$, as it should be.

The function $j(\lambda)$ can be used to show that $j(\lambda) = j(\lambda')$ iff λ' is on the S_3 orbit given by Eq. (13). It is

enough to check this for the two generators of S_3 , i.e., $\lambda \rightarrow 1/\lambda$ and $\lambda \rightarrow 1 - \lambda$.

It follows that the different values λ' arise from modular transformations in the coset $\Gamma(1)/\Gamma(2)$. This is useful, since we also know that the coset is of the form

$$S_3 \cong \langle S, T | S^2 = ST^2S = T^2 = (ST)^3 = I \rangle. \quad (14)$$

Combining this with our knowledge of the generators of the subgroups between $\Gamma(1)$ and $\Gamma(2)$, we can identify that the order 2 elements in Eq. (13) correspond to the elements S^2, T and STS of order 2 in the coset, Adding any of these to $\Gamma(2)$ will generate one of the different index 2 subgroups. Similarly, the order 3 elements in Eq. (13) correspond to ST and TS , but if we add one of these we have the other one as well, since $(ST)^2 = TS$ in the coset. Adding these transformations to $\Gamma(2)$ gives the subgroup Γ_2 of index 2.

The above also leads to the observation that fixed points of λ under the action (13) correspond to special curves \mathcal{E}_λ with enhanced modular symmetries generating the subgroups between $\Gamma(1)$ and $\Gamma(2)$. Therefore the λ -line is no longer 6-to-1 over the j -line but less, and the corresponding fixed points can be found in Fig. 3.

We have now explained how the $\Gamma(2)$ -modular invariant λ , along with its symmetries and relations to the subgroups at level 2, arise from the family of level two curves \mathcal{E}_λ . From these facts it is possible to construct invariant rational functions $f(\lambda)$ for the subgroups at level 2. However, to fix which modular transformation is which in Eq. (13), we need to know precisely how the torsion points e_i map to the lattice \mathbb{C}/Λ_τ . This can be done using theta functions, and will lead to the most useful explicit representation of λ .

III. MODULAR CURVES

In order to find explicit functional forms for the highly transcendental functions discussed above, we need a better understanding of the level 2 moduli space, i.e., the modular curve $\mathcal{M}(2)$ parametrized by “the λ -line”. This will give us access to a supply of level 2 modular forms for $\Gamma(2)$ and the other level 2 subgroups. As a by-product we can also pinpoint the modular transformations in the coset $\Gamma(1)/\Gamma(2)$ acting on λ .

A. Level 2 moduli and theta functions

We explicitly construct the branched covering of \mathcal{E}_τ over \mathbb{CP}^1 closely following Ref. 16. The functions $\theta_4(z, \tau)^2$ and $\theta_1(z, \tau)^2$ on \mathcal{E}_τ satisfy

$$\begin{aligned} f(z + 1, \tau) &= f(z, \tau) \\ f(z + \tau, \tau) &= e^{-2\pi i \tau} e^{-4\pi i z} f(z, \tau). \end{aligned}$$

In other words, they are quasi-doubly-periodic in z with Fourier expansions in z . It follows that the Fourier coefficient c_{n+2} of f is determined by c_n . The space of holomorphic functions satisfying the above quasi-periodicity is therefore at most two dimensional, and in fact spanned by $\theta_4(z, \tau)^2$ and $\theta_1(z, \tau)^2$, since they are linearly independent.

There is a map

$$h : E_\tau \cong \mathbb{C}/\Lambda_\tau \rightarrow \mathbb{CP}^1, \quad z \mapsto (\theta_4(z, \tau)^2, \theta_1(z, \tau)^2),$$

which is well defined due to the transformation properties of the theta functions. This is a degree $d = 2$ map, so the Riemann-Hurwitz theorem for $h : X \rightarrow Y$ gives

$$2g_X - 2 = d(2g_Y - 2) + \sum_P (e_P - 1).$$

Since $e_P \leq \deg h$, the map must be ramified along four points with $e_P = 2$. Furthermore, because $\theta_4(\tau/2, \tau) = 0$ and $\theta_1(0, \tau) = 0$ are the only zeros, two of the branch points are

$$h(\tau/2) \sim (0, 1), \quad h(0) \sim (1, 0).$$

By the lifting property, the following map has a lift from

$$\mathbb{CP}^1 \rightarrow \mathcal{E}_\tau, \quad p \mapsto \sum_{P \in h^{-1}(p)} e_P P.$$

to an entire map on the universal cover \mathbb{C} of \mathcal{E}_τ (or \mathbb{CP}^1) and therefore by Liouville's theorem is a constant. Since $2 \cdot 0 = 2 \cdot (\tau/2) = 0$ on \mathcal{E}_τ , we therefore also have for the other two ramification points P_i that

$$2P_i = 0 \pmod{\Lambda_\tau}.$$

These branch points are therefore

$$\begin{aligned} &(\theta_4(1/2, \tau)^2, \theta_1(1/2, \tau)^2) \\ &(\theta_4(1/2 + \tau/2, \tau)^2, \theta_1(1/2 + \tau/2, \tau)^2). \end{aligned}$$

It is convenient to express these theta functions in terms of the Jacobi constants $\theta_i(0, \tau)$:

$$\begin{aligned} \theta_1(1/2, \tau) &= -\theta_2(0, \tau) \\ \theta_4(1/2, \tau) &= \theta_3(0, \tau) \\ \theta_4(1/2 + \tau/2, \tau) &= e^{-\pi i \tau/4} \theta_2(0, \tau) \\ \theta_1(1/2 + \tau/2, \tau) &= e^{-\pi i \tau/4} \theta_3(0, \tau), \end{aligned}$$

which gives the four branch points in the form

$$(1, 0), \quad (0, 1) \tag{15}$$

$$\left(\frac{\theta_3(0, \tau)^2}{\theta_2(0, \tau)^2}, 1 \right), \quad \left(\frac{\theta_2(0, \tau)^2}{\theta_3(0, \tau)^2}, 1 \right). \tag{16}$$

We see that the branch points are $0, 1, i\infty$ and

$$\lambda(\tau) = \frac{\theta_2(\tau)^4}{\theta_3(\tau)^4}, \tag{17}$$

in affine coordinates $x \in \mathbb{CP}^1$. Since the branch points are the images of 2-torsion points of \mathcal{E}_τ , it follows that the cubic equation of the elliptic curve is given by Eq. (12), and (17) gives an explicit functional form for λ whose modular properties are transparent. We have therefore reconstructed the family of level 2 elliptic curves in Legendre form, while obtaining a very useful explicit form for λ in terms of Jacobi theta-functions.

Observe that our choice of theta functions led to the non-trivial branch point λ , and other combinations of Jacobi theta function (see the Appendix) give the other values on the S_3 orbit. These choices also correspond to the permutation of the roots $e_1 = 0$, $e_2 = 1$ and $e_3 = \lambda$ of the cubic equation, or equivalently the branch points of the cover $\mathbb{CP}^1 \rightarrow \mathcal{E}_\tau$. Since we have four points $0, 1, \lambda$ and ∞ on the Riemann sphere \mathbb{CP}^1 , their cross ratio is an invariant under the automorphisms of \mathbb{CP}^1 . We can only reach branch points different from λ by permuting the points in the cross ratio. This is the action of S_3 on λ .

B. Modular functions at level 2

Having analyzed elliptic curves and their equations at level 2 in considerable detail, we are ready to present the modular functions associated to these moduli spaces.

A meromorphic function on \mathbb{H} that transforms covariantly or contravariantly under the (sub-)modular group is called a *modular function*. These are building blocks of automorphic RG potentials and β -functions that embody the constraints necessary for any RG flow consistent with the dualities.

For the group $\Gamma(2)$, we have already met an invariant function (or a hauptmodul) in Eq. (17). We have

$$\begin{aligned} \lambda(0) &= 1, & \lambda(1) &= \infty, & \lambda(i\infty) &= 0, \\ \lambda'(0) &= 0, & \lambda'(1) &= \infty, & \lambda'(i\infty) &= 0. \end{aligned}$$

λ is a univalent invariant function, so it takes every complex value once in the (proper) fundamental domain of $\Gamma(2)$ and gives an explicit map $\lambda : \mathbb{H}/\Gamma(2) \rightarrow \mathbb{CP}^1$.

From the transformation properties of the theta functions (Appendix) it follows that the $\Gamma(2)$ -invariant function λ transforms in the representation

$$\begin{aligned} \lambda|_S &= 1 - \lambda, & \lambda|_T &= \frac{\lambda}{\lambda - 1}, & \lambda|_{STS} &= \frac{1}{\lambda} \\ \lambda|_{ST} &= \frac{1}{1 - \lambda}, & \lambda|_{TS} &= \frac{\lambda - 1}{\lambda} \end{aligned} \tag{18}$$

under modular transformations in the coset $S_3 \cong \Gamma(1)/\Gamma(2)$, and $\lambda|_\gamma(\tau) := \lambda(\gamma(\tau))$. This is the S_3 orbit discussed above (Eq. (13)), and is isomorphic to the one given by Eq. (32); $S_3 \cong \Gamma(1)/\Gamma(2)$ acts as automorphisms¹⁵ of $\mathcal{M}(2) := \mathbb{H}/\Gamma(2) \cong \mathbb{CP}^1$. Eq. (18) gives an explicit formula for the transformation properties of the derivative λ' , but it is often more convenient

to work with combinations of λ and the theta functions (Appendix), since they are more familiar.

The hauptmoduli for the other groups $\Gamma_0(2)$, $\Gamma^0(2)$ and $\Gamma_\theta(2)$ can all be written as rational functions of λ :

$$\begin{aligned}\lambda_T &\equiv \lambda_0 = \frac{\lambda - 1}{\lambda^2} \\ \lambda_R &\equiv \lambda^0 = \lambda(1 - \lambda) \\ \lambda_S &\equiv \lambda_\theta = -\frac{\lambda}{(1 - \lambda)^2}.\end{aligned}$$

It follows from Eq. (18) that they are invariant under the corresponding groups.

All groups mentioned here have genus-0 modular curves $\mathcal{M}(\Gamma) := \mathbb{H}/\Gamma$, which in particular means that there is a one-to-one map λ_Γ to the Riemann sphere \mathbb{CP}^1 , see Ref. 15, generalizing the j -invariant for $\Gamma(1)$. All meromorphic modular forms of weight 0 for Γ are rational functions of λ_Γ , and we call this function field $M_0(\Gamma) = \mathbb{C}(\lambda_\Gamma)$.

Similarly as with invariant functions for the group Γ , one can define functions that transform covariantly or contravariantly under modular transformations

$$f(\gamma(\sigma)) = (c\sigma + d)^k f(\sigma), \quad \gamma \in \Gamma, \quad \sigma \in \overline{\mathbb{H}}$$

and can be thought of as tensors under modular transformations. Here the weight k is even, since $-I \in \Gamma$, so that odd forms vanish identically. We also define the notation $f_k|_\gamma := (c\tau + d)^{-k} f_k(\gamma(\tau))$, generalizing the earlier one.

The space of *holomorphic* functions with non-zero positive weight on $\overline{\mathbb{H}}$ is highly constrained, much like the space of invariant meromorphic functions (the weightless case), and in fact always finite dimensional. Clearly the space of weight k modular forms $M_k(\Gamma)$ for Γ is a vector space, whose dimensions¹⁵ for $k = 0$ and $k = 2$ are shown in Table I. There is no weight 2 modular form for $\Gamma(1)$.

$\dim M_k(\Gamma)$	$\Gamma(1)$	Γ_2	$\Gamma_0(2)$	$\Gamma^0(2)$	$\Gamma_\theta(2)$	$\Gamma(2)$
$k = 0$	1	1	1	1	1	1
$k = 2$	0	0	1	1	1	2

Table I: Dimensions of the vector spaces $M_k(\Gamma)$ ($k = 0, 2$) for level 2 groups.

We have already met the weight 4 and 6 forms for $\Gamma(1)$. In fact, these generate all modular forms for $\Gamma(1)$ ¹⁵.

We construct the generators of weight 2 forms for the subgroups that are relevant to the QHE. Since λ is a ratio of two theta functions to the fourth power, one can suspect that they have something to do with modular forms for $\Gamma(2)$. In fact, using the transformation properties of the theta functions, one see that

$$\theta_2(\tau)^4, \quad \theta_3(\tau)^4, \quad \theta_4(\tau)^4 \quad (19)$$

are all weight 2 forms for $\Gamma(2)$. This space is only two-dimensional and the linear relation between them is the Jacobi identity $\theta_3^4 = \theta_2^4 + \theta_4^4$.

Since the forms above $\Gamma(2)$ are contained in $M_2(\Gamma(2))$, these will be linear combinations of the generators θ_2^4 and θ_3^4 . Inspection shows that

$$\beta^0(\tau) = \theta_3^4 + \theta_2^4 \quad (20)$$

$$\beta_0(\tau) = \theta_3^4 - \frac{\theta_2^4}{2} \quad (21)$$

$$\beta_\theta(\tau) = \theta_3^4 - 2\theta_2^4 \quad (22)$$

are covariant vector fields for the index 3 subgroups. Modular transformations in $\Gamma(1)/\Gamma(2)$ “permute” these forms much like for the λ . Explicitly, $\beta^0|_T = \beta_\theta$, $\beta_\theta|_T = \beta^0$, $\beta^0|_S = -2\beta_0$ and $\beta_0|_S = -\beta^0/2$; but it follows that $\beta_\theta + \beta^0 - 2\beta_0 = 0$, which is necessary since no weight 2 form exists for $\Gamma(1)$. We have now found all the generators for weightless functions and weight 2 forms for level 2 symmetries. As shown in Ref. 2, all these arise from a family of RG potentials Φ_a as $\beta_a = -\partial_\sigma \Phi_a \propto \theta_3^4 - a\theta_2^4$.

In general, the construction of modular forms may seem to be a daunting task, but the following two results¹⁵ greatly facilitate the job. First, let λ_Γ be a univalent¹⁵ weightless modular form for a genus-0 subgroup $\Gamma \subset \Gamma(1)$ (all cases considered here are univalent since the functions have first order pole at one of the inequivalent cusps). If λ_q is a modular form with a q th order pole at $\mathbb{F} - \mathbb{E}$, then $\lambda_q = P(\lambda_\Gamma)/Q(\lambda_\Gamma)$ and

$$\beta_{2k}(\lambda_\Gamma) = \frac{R(\lambda_\Gamma)}{S(\lambda_\Gamma)} \lambda_\Gamma'^k$$

is a weight $2k$ modular form. In other words, $\beta_{2k} \in M_{2k}(\Gamma)$, where P , Q , R , and S are polynomials and the degrees of P and Q do not exceed q . In particular, the (logarithmic) derivative of a weightless modular form is a modular form of weight 2. Note, however, that there may be modular forms that are not derivatives of modular functions.

Second, let f be a modular form of weight $2k$ for the group $\Gamma \subset \Gamma(1)$ of index μ . Then the “sum rule”

$$n_\infty + \sum_{z_c = \gamma(i\infty), \gamma \notin \Gamma} n_c + \sum_{\otimes} \frac{1}{|\Gamma_{z_\otimes}|} n_\otimes + \sum_p n_p = \frac{\mu k}{6}$$

follows from an application of Cauchy’s theorem to the fundamental domain. Here n_P is the leading power of $f(z) \sim a_p(z - z_p)^{n_p} + \dots$ around a pole or zero z_p , n_\otimes is the same quantity for a fixed point $z_\otimes \in \mathbb{E}$ of Γ , $|\Gamma_{z_\otimes}|$ is the order of the isotropy group of z_\otimes , and n_∞ is the leading power of the q -expansion of $f(z)$ around the cusp $z = i\infty$.

For a level N form $f(\gamma(z)) = (cz + d)^{2k} f(z)$ $\gamma \in \Gamma(N)$, there are several inequivalent cusps. The order n_c of f at a cusp $z_c = \gamma_0(i\infty)$ is defined as the leading order of the $q_N = e^{2\pi iz/N}$ expansion of $f|_{\gamma_0}(z)$ at $z = i\infty$. For example, at level 2 n_0 is the leading power of $z^{-2k} f(-1/z)$ expanded around $z = i\infty$ in powers of $q^{1/2}(-z) = e^{-i\pi z}$.

For the subgroups of index 3 or 6 this reduces to

$$n_\infty + n_0 + \frac{1}{2}n_\otimes + \sum_p n_p = \frac{k}{2},$$

or equivalently

$$n_\infty + n_0 + n_1 + \sum_p n_p = k.$$

These theorems are useful since they directly constrain the spaces of modular forms and functions based on the analytic structure. With a bit more work, upper bounds on the dimensions¹⁵ of the relevant spaces M_{2k} cited above can be constructed — in particular, they are finite.

As explained at length in Ref. 2 for the QHE, this sum rule classifies the fixed points (zeros) or instabilities (poles) of the potentials and β -functions, which correspond to the fixed points and singularities of the effective field theories with the given symmetry.

Using the defining q -series and transformation properties of the theta functions, we can see that

	θ_2^4	θ_3^4	θ_4^4	
n_∞	1	0	0	
n_1	0	1	0	
n_0	0	0	1	

In particular, a linear combination $\beta_2 \sim a\theta_2^4 + b\theta_3^4 \in M_2(\Gamma(2))$ will have a simple zero somewhere on the upper halfplane \mathbb{H} .

C. Effective field theories

The following discussion applies to any quantum field theory with modular symmetry, but we now start to focus on RG fixed points and critical points of the quantum Hall system. These points belong to moduli spaces that are modular curves associated with the elliptic curves considered above. More details can find be found in the Appendix, and a comprehensive account of the symmetries and their representations is given in our companion paper².

Phenomenology indicates that a modular group Γ acts as dualities in the parameter space of the QHE and the quantum critical and fixed points are special points of the modular curve $\mathcal{M}(\Gamma) = \overline{\mathbb{H}}/\Gamma$, where the compactification $\overline{\mathbb{H}}$ of \mathbb{H} is the physical parameter space. $\mathcal{M}(\Gamma)$ parametrizes elliptic curves with symmetry Γ , and since any $\Gamma \subset \Gamma(1)$ acts properly discontinuously on $\overline{\mathbb{H}}$ it is a Riemann surface. For a discussion suitable for our application, see for example Ref. 14.

The precise form of the automorphism group Γ has fundamental consequences for the global phase diagram of the QHE. Since the quantum Hall plateaux, i.e., the attractive (real) fixed points (\oplus), are always observed at rational values of the Hall conductance $\sigma = \sigma_H = p/q \in \mathbb{Q}$, they are related to number-theoretic properties of the modular group. Filling fractions where no plateau is observed are repulsive rational fixed points (\ominus) in this

model. All of these follow from the repulsive “metallic” phase at $\sigma = i\infty$, the filled Landau level $\sigma = 1$, and the attractive quantum Hall insulator at $\sigma = 0$ by modular transformations, and the set of attractive and repulsive fixed points make up the boundary $\partial\overline{\mathbb{H}} = \mathbb{Q} \cup \{i\infty\}$ of the compactified upper half space $\overline{\mathbb{H}}$. This follows since the fixed point set $\{\oplus\} \cup \{\ominus\}$ (see Fig. 1 (d)) maps to itself under modular transformations

$$\gamma(\infty) = \frac{a}{c}, \quad \gamma(0) = \frac{b}{d}, \quad \gamma(1) = \frac{a+b}{c+d},$$

and the plateaux-structure in the QHE follows and implies modular symmetries, as detailed in the Appendix.

To these boundary fixed points we must add possible RG fixed points inside \mathbb{H} . If we assume that fixed points of Γ are the only RG fixed points, then the set of inequivalent RG fixed points is given by the set inequivalent fixed points of Γ . These have the following classification for the full modular group $\Gamma(1)$.¹⁵

1. Elliptic fixed points of order 2 (\otimes). Since these points are fixed by a transformation that squares to one it must be conjugate to S , and the set of elliptic points of order 2 is

$$\mathbb{E}_2 = \{\sigma \in \overline{\mathbb{H}} | \sigma = \gamma^{-1}(i), \gamma \in \Gamma(1)\}.$$

The isotropy group of any point in \mathbb{E}_2 is \mathbb{Z}_2 .

2. Elliptic fixed points of order 3. Since these points are fixed by a transformation that cubes to one it must be conjugate to ST or TS , and the set of elliptic points of order 3 is

$$\mathbb{E}_3 = \{\sigma \in \overline{\mathbb{H}} | \sigma = \gamma^{-1}(\omega), \gamma \in \Gamma(1)\},$$

with $\omega = e^{2\pi i/3}$. The isotropy group of any point in \mathbb{E}_3 is \mathbb{Z}_3 . We will also use the notation $\mathbb{E} = \mathbb{E}_2 \cup \mathbb{E}_3$.

3. Parabolic fixed points or *cusps*. Since these points are fixed by a transformation conjugate to T^n ($n \in \mathbb{Z}$), the set of parabolic fixed points is

$$\mathbb{P} = \{\sigma \in \overline{\mathbb{H}} | \sigma = \gamma^{-1}(\infty), \gamma \in \Gamma(1)\}.$$

There are also hyperbolic fixed points that are fixed by transformations conjugate to transformations of infinite order, but this is the set of irrational numbers¹⁵ and does not concern us.

The classifications for subgroups $\Gamma \subset \Gamma(1)$ are analogous, but some fixed point sets may be empty if there are no transformations in Γ of the type listed above. In particular, the subgroups at level 2 of index three only have \mathbb{E}_2 points and the subgroup of index two only points \mathbb{E}_3 , in addition to the cusps. The fixed point sets that do appear are partitioned into to equivalence classes mod Γ .

The interpretation of these fixed points in terms of elliptic curves is the following. The cusps \mathbb{P} represent degenerate elliptic curves at the boundary of the moduli

space, whereas the elliptic points \mathbb{E} are curves with “extra” symmetries or automorphisms induced by the fixed-point transformations $\gamma \in \Gamma$.

Locally, near any fixed point in \mathbb{E}_2 , \mathbb{E}_3 or \mathbb{P} , the flow looks the same, since the points are related by an element $\gamma \in \Gamma$ that commutes with the flow. Around a point in \mathbb{E}_2 , the flow bifurcates, with one irrelevant and one relevant direction. Similarly, the flow around a point in \mathbb{E}_3 “trifurcates”.

Consider now fundamental domains. By definition, a (proper) fundamental domain \mathbb{F} is defined as the set in \mathbb{H} so that no two points in \mathbb{F} are equivalent under Γ . We will loosen this definition so that within a fundamental domain, two points on the boundary may be equivalent, i.e., we take the closure of the proper fundamental domain. The topology (fixed point structure) of the RG flow is therefore indeed captured by the modular curve $\mathcal{M}(\Gamma) = \mathbb{H}/\Gamma$, which is in fact a compact Riemann surface, since Γ acts properly discontinuously on \mathbb{H} . Considered as a set in \mathbb{H} , \mathbb{H}/Γ is the fundamental domain \mathbb{F}_Γ of Γ . It is easy to see that any fixed point of Γ is on the boundary of \mathbb{F} .¹⁵ Therefore, with the assumptions made, the set of critical points is on the boundary of the fundamental domain. Moreover, the flow cannot cross the boundaries of the fundamental domain, and therefore the boundary segments of $\partial\mathbb{F}$ and their images act as separatrices for the flow. These boundary segments are also geodesics of the hyperbolic metric on \mathbb{H} . The covariant β -function will be a weight two modular form on $\mathcal{M}(\Gamma)$.

We have constructed² a family of Γ_a invariant RG flows, with $\Gamma(2) \subseteq \Gamma_a \subset \Gamma(1)$ parametrized by $a \in \mathbb{R}$, so that the critical point \otimes is on the boundary of the fundamental domain \mathbb{F}_2 of $\Gamma(2)$. For a general value of a , this critical point \otimes on \mathbb{H} is not fixed by any element of Γ_a . However, the critical point is inherited from a bigger group $\Gamma_X \supset \Gamma(2)$ ($X = S, T, R$), which appear for special values a_R , a_T , and a_S of a , and in these cases the critical point on \mathbb{H} is fixed by an element in Γ_X . Otherwise its location moves continuously along the boundaries of the fundamental domain of the submodular group $\Gamma(2)$ as the parameter $a \in \mathbb{R}$ is varied. The set of attractive $\{\oplus\}$ and repulsive $\{\ominus\}$ points is entirely determined by the cusps of the subgroups Γ_X and this highly constraints the topology of the global RG flow.

IV. MODULAR RG FLOWS IN THE QHE

Having assembled all the mathematical tools required, we can now build modular models that contain the universal data and RG flows for quantum Hall systems. This will lead us to a 1-parameter family of curves. The geometry of this family appears to encode almost all available scaling data^{1,2}.

A. Renormalization

An RG flow is a vector field on the space of those parameters that are relevant at the chosen energy scale. The QHE is parametrized by the conductivity tensor σ^{ij} , or equivalently its inverse, the resistivity matrix $\rho^{ij} = (\sigma^{-1})^{ij}$. These transport tensors are non-trivial because the background magnetic field breaks parity (time-reversal) invariance, whence the off-diagonal Hall coefficient is permitted by the generalized Onsager relation. A conventional Hall bar exhibiting fractional plateaux is usually a layered structure that confines all charge transport to a single isotropic layer of size $L_x \times L_y$, aligned with a current I in the x -direction so that $R_{*x} = (L_*/L_y)\rho^{*x}$. The Hall resistance $R_H = \rho_H = \rho^{yx}$ is quantized in the fundamental unit of resistance, h/e^2 , while the dissipative resistance $R_D = \rho^{xx}/\square$ is rescaled by the aspect ratio $\square = L_y/L_x$.

We choose to label the low-energy parameter space by the complexified resistivity $\rho = \rho^{xy} + i\rho^{xx} = -\rho_H + i\rho_D$, or equivalently by the complexified conductivity $\sigma = \sigma^{xy} + i\sigma^{xx} = \sigma_H + i\sigma_D$. These complex coordinates only take values in the upper half of the complex plane, $\mathbb{H} = \{\sigma \in \mathbb{C} | \text{Im } \sigma = \sigma_D > 0\}$, because the dissipative conductivity (resistivity) is positive. The reason for not including the real line in our definition of the parameter space will soon become clear.

The components of the tangent vectors $\beta^\sigma = \beta^1 + i\beta^2$ to the RG flow on the space of conductivities are the Gellman-Low β -functions, which describe the (logarithmic) rate of change of the renormalized couplings with respect to the scale parameter Λ :

$$\beta^1 = \frac{d\sigma_H}{dt}, \quad \beta^2 = \frac{d\sigma_D}{dt} \quad \text{with } t = \ln(\Lambda/\Lambda_0). \quad (23)$$

The flow ends at stable fixed points, which are the plateaux observed at rational values $\sigma_\oplus = \sigma_H \in \mathbb{Q}$ of the complex conductivity, so these are the only real values that should be included in the physical parameter space. The quantum critical points $\sigma_\otimes \in \mathbb{H}$ are saddle points of the flow, so the β -function must vanish at these points, and only at these points. Expanding the β -function around a critical point gives universal linear coefficients $1/\nu_i$, and the real numbers ν_i are the critical exponents.

Careful examination^{1,2} of the fixed point structure of scaling data have revealed that there appear to be emergent symmetry groups $\Gamma \subset \text{Aut}\Gamma(1) = \text{AutPSL}(2, \mathbb{Z})$, as discussed in sect. III C. A first indication of this is that compactifying the topology of the spaces on which the modular groups act recovers the plateaux values needed to complete the physical parameter space. For the full modular group, for example, the space is $\mathbb{H} = \mathbb{H} \cup \mathbb{Q} \cup \{i\infty\}$, but is further partitioned into attractive (\oplus) and repulsive (\ominus) plateaux along the rationals for the subgroups.

This is a very strong statement, with three immediate

and unavoidable implications for phenomenology (low-energy data) that at first sight appear far too rigid to be of use in physics. However, it is precisely this rigidity that gives Γ -symmetry teeth, and which allows it to be easily identified.

First, as discussed in the previous section, any Γ -symmetry partitions parameter space into *universality classes* that are suitable images of $\mathcal{M}(\Gamma) \cong \mathbb{H}/\Gamma$, with each phase “attached” to a unique plateau fixed point on the real line, see Fig. 1 (d). This is not like a gauge symmetry, which identifies gauge equivalent field configurations. Rather, it collects or classifies effective field theories into families (universality classes) that are related, but not identified, under modular transformations. The fixed point set of each Γ cannot be manipulated: all the plateaux and quantum critical points following from Γ must be included. The quantum critical points are located on the phase boundaries, and as soon as one is pinned down the location of all the others is fixed.

Secondly, since the actions of the renormalization group and the modular group must be consistent, Γ forces the RG flow into a straight-jacket. Not only are the RG fixed points, including the quantum critical (saddle) points, mapped into each other, so are the β -functions, and it follows that the critical exponents are “super-universal”: they are always the same, independent of which quantum phase transition is considered. Furthermore, given an initial value for a flow the geometry of the entire flow-line is largely fixed by the symmetry. The flows observed experimentally in the spin-polarized QHE appear to satisfy this prediction as well.

Thirdly, it is easy to see that if the β -function is a holomorphic (complex analytic) function of the complex parameter, then the critical exponents can not have opposite signs, i.e., the critical point is not a saddle point, which would be in contradiction with experiment. Remarkably, it is a mathematical fact that no modular holomorphic β -functions exist, so this cannot happen. However, non-holomorphic modular forms are very difficult to handle, and explicit examples are all but non-existent. Fortunately, it may not be necessary to abandon analyticity altogether. The covariant β -function $\beta_\sigma = g_{\sigma\bar{\sigma}}\beta^{\bar{\sigma}}$ can be modular and holomorphic. This quarantines non-holomorphicity to the metric g , a modular form of weight $(2, 2)$. It is non-singular at the critical points, and therefore invertible, unless the effective field theory breaks down at these points. We have no reason to believe that it will, and this is verified *a posteriori* by the agreement of scaling data with the resulting form for the β -function:

$$\beta^{\text{phys}} = \beta^\sigma = g^{\sigma\bar{\sigma}}\beta_{\bar{\sigma}} = g^{\sigma\bar{\sigma}}\partial_{\bar{\sigma}}\Phi. \quad (24)$$

This has the consequence that $\nu_1 = -\nu_2$, a phenomenon dubbed “anti-holomorphic scaling” in Ref. 1. The fact hypothesized in Ref. 1 and explicitly verified in Ref. 17, that modular covariant β -functions are gradients of a “holomorphically factorized” RG potential (this is a slight abuse of notation, since it is $\exp\Phi(\sigma, \bar{\sigma}) =$

$\varphi(\sigma)\bar{\varphi}(\bar{\sigma})$ that is holomorphically factorized):

$$\Phi(\sigma, \bar{\sigma}) = \ln \varphi(\sigma) + \ln \bar{\varphi}(\bar{\sigma})$$

is physically reasonable and therefore lends further support to the idea of anti-holomorphic scaling. It is also a very useful simplification, since it lifts the geometric analysis of renormalization from vector to scalar fields.

B. RG potentials

Our main goal in Ref. 2 was to construct an RG potential interpolating between all the level 2 symmetries and compare it with available experimental data. We now briefly recollect it. The potential is a Γ -modular function Φ whose gradient is the covariant β -function. Φ is a real-valued modular function that is finite in the finite part of the upper half plane, i.e., what we are calling a physical potential. To this reasonable list of constraints we add that the potential should be holomorphically factorized. This is certainly the case at the conformal fixed points, and agrees with the flows observed in the spin-polarized case. Together these conditions guarantee the existence of a covariant holomorphic β -function that vanishes at the quantum critical points.

The three congruence subgroups Γ_i ($i = 1, 2, 3$) at level two admit physical potentials $\varphi_i = \ln f_i$ ($i = 1, 2, 3$), where the f_i are elementary modular functions (essentially ratios of theta-functions) defined in the Appendix. Since they all contain $\Gamma(2)$ as their largest subgroup, we consider the most general maximally symmetric ($\Gamma(2)$ -invariant) form for the RG potential, i.e., $\Phi = \psi + c.c.$, where ψ is the linear superposition of elementary potentials:

$$\psi = c_1\varphi_1 + c_2\varphi_2 + c_3\varphi_3 \quad (c_i \in \mathbb{C}).$$

At first sight this ansatz looks useless, since it requires six real parameters. However, for Φ to factorize holomorphically the coefficients c_i must be real, giving

$$\Phi(\sigma, \bar{\sigma}) = \ln |f_1|^{2c_1} + \ln |f_2|^{2c_2} + \ln |f_3|^{2c_3}.$$

Up to an arbitrary normalization factor that only affects flow rates, not their shapes, this leaves only two real parameters. Furthermore, since it turns out that all the f_i are simple fractions of polynomials in the $\Gamma(2)$ -modular function λ , the maximally symmetric family of quantum Hall RG potentials reduces to

$$\Phi_a \propto \ln \lambda(\lambda - 1)^{a-1} + c.c. .$$

This gives the covariant β -function

$$\beta_a(\sigma) = -\partial_\sigma \Phi_a \propto \theta_3^4(1 - a\lambda) = \theta_3^4 - a\theta_2^4,$$

which vanishes as it should at the quantum critical points $\sigma_\otimes(a) = \lambda^{-1}(1/a)$. In particular, we can invert this relation

$$a(\sigma_\otimes) = \frac{1}{\lambda(\sigma_\otimes)}. \quad (25)$$

The $\Gamma(2)$ -symmetry of the potential Φ_a along with the covariant β_a -function is enhanced when a takes certain integer values: to Γ_R when $a_R = -1$, to Γ_T when $a_T = 1/2$, and to Γ_S when $a_S = 2$. These forms at level 2 were discussed above.

Almost all scaling data appear to be consistent with the RG potential Φ_a for various values of a , not restricted to the enhanced symmetry points $a = a_i$, but always real. The interpolating forms β_a for $a \in \mathbb{R}$ between the level 2 subgroups are shown in Fig. 4.

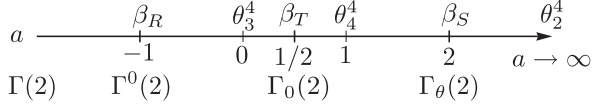


Figure 4: The vector field β_a is $\Gamma(2)$ symmetric for any value of the parameter $a \in \mathbb{C}$. When $a = -1, 1/2, 2$ the symmetry is enhanced to an index 3 group, and when $a = 0, 1, \infty$ the β -function degenerates to a θ -function.

C. Quantum Hall curves

We can make use of the level 2 curves that we have explained. The relation (25) can be considered to describe as the non-trivial branch point of level 2 elliptic curve in Legendre form. This immediately allows us to interpret all the universal properties of the QH system exhibiting the scaling with Φ_a as originating from the elliptic curve

$$\mathcal{E}_a : y^2 = x(x-1)(x-a). \quad (26)$$

As long as the value of a is not a fixed point under the S_3 action on $a = 1/\lambda$, this QH system will have a $\Gamma(2)$ -symmetry. For fixed points, the $\Gamma(2)$ symmetry is enhanced to one of the subgroups at level 2. For real values of a , only the index 3 groups can appear, and in between the enhanced symmetries the flow satisfies the generalized semi-circle law as explained in Ref. 2.

The non-trivial branch point a describes the critical saddle-point σ_\otimes and the other branch points $0, 1, i\infty$ correspond to filled Landau level $\sigma = 1$, the QH insulator $\sigma = 0$, and the repulsive “metallic” point $\sigma = i\infty$, respectively, which are the special points at the boundary of the parameter space.

Finally, the weight 2 modular form $\beta_a = -\partial_\sigma \Phi_a = \theta_3^4 - a\theta_2^4 \in M_2(\Gamma(2))$, unique up to real normalization, generates the RG flow of the system, and by construction has a saddle-point zero at $\sigma_\otimes(a)$.

As we have shown in Ref. 2, changing the value of $a \in \mathbb{R}$ continuously interpolates between the different subgroups, apart from discontinuities at $a = 0, 1, i\infty$, where the curve degenerates, which are necessary to exist between the enhanced-symmetry curves at the values $a = -1, 1/2, 2$, which all have different sets of attractive (\oplus) and repulsive (\ominus) plateaux in their moduli.

In summary, the family of RG flows in Ref. 2 and its phenomenological implications arise from the family of elliptic curves \mathcal{E}_a .

D. Modular action on the parameter a

Finally, we show how to deduce the interpretation of the parameter a given above without using the location of the critical point $\sigma_\otimes(a)$.

The RG potential Φ_a and covariant β_a -function are $\Gamma(2)$ symmetric for any value $a \in \mathbb{C}$. Under modular transformations in the coset $S_3 \cong \Gamma(1)/\Gamma(2)$, the family $\{\Phi_a, \beta_a\}$ transforms to itself and we can compute the corresponding modular action on the parameter a .

The generators S and T act on the parameter a via induced actions S^* and T^* (i.e., as “pullbacks”), in the sense that for any $\gamma \in \Gamma(1)$, one has

$$\begin{aligned} \Phi_a(\gamma(\sigma)) &= \nu(\gamma, a) \Phi_{\gamma^*a}(\sigma) + \text{constant} \\ \beta_a(\gamma(\sigma)) &= f(\sigma, a|\gamma) \beta_{\gamma^*a}(\sigma). \end{aligned}$$

One can see that $\beta_{\gamma^*a}(\sigma)$ should be a modular form for $\gamma^{-1}\Gamma_a\gamma$, if β_a is a form for Γ_a . Naturally, $\gamma^*a = a$ for any $\gamma \in \Gamma(2)$, so in reality the action is non-trivial for $\gamma \in \Gamma(1)/\Gamma(2)$. The action γ^* inherits the group law from the action of $\Gamma(1)$ on \mathbb{H} .

The multiplier factor $f(z, a|\gamma)$ is the modular weight $(\gamma : z)^{2k} := (cz + d)^{2k}$ for $\gamma \in \Gamma_a$. Otherwise it differs and in general has an a -dependent factor. We have the consistency conditions:

$$\begin{aligned} \nu(\gamma_1, \gamma_2^*a) \nu(\gamma_2, a) &= \nu(\gamma_2\gamma_1, a) \\ f(\gamma_1(\sigma), a|\gamma_1) f(\sigma, \gamma_2^*a|\gamma_1) &= f(\sigma, a|\gamma_2\gamma_1). \end{aligned}$$

Both the actions of S^*, T^* and the multipliers $\nu(\gamma, a)$, $f(\sigma, a|\gamma)$ can be deduced from the known modular properties of λ and theta functions. From

$$\theta_3^4(a\lambda - 1)|_T = \theta_3^4((1-a)\lambda - 1) \quad (27)$$

$$\theta_3^4(a\lambda - 1)|_S = (a-1)z^2 \theta_3^4(a\lambda/(a-1) - 1) \quad (28)$$

we get

$$\begin{aligned} \nu(T, a) &= 1, & \nu(S, a) &= a - 1 \\ f(\sigma, a|T) &= 1, & f(\sigma, a|S) &= (a-1)z^2. \end{aligned}$$

So we see that $f(\sigma, a|\gamma) = \nu(\gamma, a)(\gamma : z)^{2k}$. Then

$$T^*a = 1 - a, \quad S^*a = \frac{a}{a-1}$$

and this specifies the action of the modular group on the parameter a . For the non-trivial transformations $\gamma \in \Gamma(1)/\Gamma(2)$ we obtain

$$a \xrightarrow{\gamma^*} \left\{ a, 1-a, \frac{1}{a}, \frac{1}{1-a}, \frac{a}{a-1}, \frac{a-1}{a} \right\}. \quad (29)$$

These transformations generate a representation of S_3 isomorphic to the one in Eq. (18) by the inversion $\lambda \rightarrow 1/\lambda$. Under modular transformations that are not in the symmetry group, the complexified β -functions get scaled by an a -dependent factor as in Eq. (28) and transforms according to the S_3 transformations in Eq. (18). This factor can be absorbed into the unknown normalization.

The points of enhanced symmetry correspond to the fixed points of S_3 under the group action (29), since then there is extra modular transformations in the group Γ_a , see Fig. 3. The fixed points of the transformations generated by S^* and T^* are

$$a = \{\pm 1, 0, 2, 1/2, -\omega^2, -\omega, i\infty\},$$

see Fig. 3. The real values $a = \{a_W, a_T, a_S\} = \{-1, 1/2, 2\}$ correspond to the groups $\Gamma^0(2)$, $\Gamma_0(2)$ and $\Gamma_\theta(2)$ that we have been discussing. The complex third roots of -1 are fixed under the transformations TS and ST . These extra symmetries generate the group Γ_2 . In this case, however, the β_a is not a modular form of weight 2, but a modular form with a non-trivial multiplier since the multipliers are complex. The multipliers in this case are ω and ω^2 , respectively. There is no holomorphic weight 2 object for Γ_2 , so this is the best we can do. Note also that $\varphi_{a=1} = \lambda$, $\varphi_{a=0} = \lambda/(\lambda - 1)$ and $\varphi_{a=i\infty} = \lambda - 1$, so in these cases, there is no enhancement of symmetry. Instead we get the canonical potential λ for $\Gamma(2)$, its translation $\lambda|_T$ and “inversion” $\lambda|_S$. These lead to the $\Gamma(2)$ forms $i\pi\theta_4^4$, $i\pi\theta_3^4$ and $i\pi\theta_2^4$, respectively.

E. Singular curves

As an example of the utility of this algebraic machinery, we analyze the UV/IR limits of quantum Hall curves, that is, the degenerating behaviour of the curves at the boundary of the moduli space that depends on the family of curves considered.

The Legendre family \mathcal{E}_λ in Eq. (12) gives a non-singular elliptic curve for every value of the parameter $\lambda \in \mathbb{CP}^1 \setminus \{0, 1, \infty\}$. The curves \mathcal{E}_0 , \mathcal{E}_1 and \mathcal{E}_∞ are singular and represent degenerations of the Hall curve at the boundary of moduli space. The corresponding complex structures are $\lambda(\tau = 0) = 1$, $\lambda(\tau = 1) = \infty$ and $\lambda(\tau = i\infty) = 0$, and clearly give rise to degenerate tori.

It is useful to write the cubic \mathcal{E}_λ in homogenous coordinates

$$\mathcal{E}_\lambda : Y^2 Z = X^3 - \frac{\lambda_1}{\lambda_0} X^2 Z - X^2 Z + \frac{\lambda_1}{\lambda_0} X Z^2.$$

The singular curves correspond to the values $\lambda \simeq (\lambda_0 : \lambda_1)$ as $(1 : 0)$, $(1 : 1)$ and $(0 : 1)$, so

$$\begin{aligned} \mathcal{E}_0 : ZY^2 - X^2(X - Z) &= 0 \\ \mathcal{E}_1 : ZY^2 - X(X - Z)^2 &= 0 \\ \mathcal{E}_\infty : ZX(X - Z) &= 0. \end{aligned}$$

\mathcal{E}_∞ is the three concurrent lines in \mathbb{CP}^2 given by $X = 0$, $Z = 0$ and $X = Z$ that intersect at $(0 : 1 : 0)$. In order to analyze \mathcal{E}_1 we make a change of variables $X \rightarrow X - Z$, giving

$$\mathcal{E}_1 : Y^2 Z = X^2(X + Z).$$

This is the nodal curve $y^2 = x^3 + x^2$ in affine coordinates, which is topologically a sphere with a double point. This is seen by considering the map $\mathbb{CP}^1 \rightarrow \mathcal{E}_1 \subset \mathbb{CP}^2$ given by

$$(U : V) \rightarrow ((V^2 - U^2)U : (V^2 - U^2)V : U^3),$$

which is surjective onto the cubic \mathcal{E}_1 . In the affine patch $Z = 0$ there is only one point $(0 : 1 : 0)$ on the curve, which is the point O at infinity. In the affine patch $Z \neq 0$ $\mathcal{E}_1 : y^2 = x^2(x + 1)$, and we set $u = Y/X = V/U$. Then $y = ux$ and $u^2 x^2 = x^2(x + 1)$ or

$$x = u^2 - 1, \quad y = u^2(u^2 - 1),$$

which is an explicit rational parametrization of the curve on the patch $Z \neq 0$. However, the map is not injective everywhere, since at the node $(0 : 0 : 1)$ (in the original coordinates $(1 : 0 : 1)$)

$$u = V/U = y/x = \pm 1,$$

which confirms that the nodal cubic is topologically a sphere with a double point. \mathcal{E}_0 is also a nodal curve. This time the rational map $\mathbb{CP}^1 \rightarrow \mathcal{E}_0 \subset \mathbb{CP}^2$ is given by

$$(U : V) \rightarrow ((V^2 + U^2)U : (V^2 + U^2)V : U^3),$$

and the double point is at $(0 : 0 : 1)$ with $u = \pm i$.

This shows that a quantum Hall curve degenerates to singular geometries at the cusps, but because these points are infinitely far away from the scaling region where the curve accounts for scaling data this does not appear to invalidate the modular model.

V. CONCLUDING REMARKS

We have shown how the modular symmetries that have been found to be consistent with most available scaling data from quantum Hall systems^{1,2} derive from a family of algebraic curves of the elliptic type.

Holomorphic modular symmetries just barely admits the existence of functions and forms that we need in order to give a quantitative description of observed RG flows, but no more. Unphysical holomorphic β -functions are prohibited by the symmetry. We have here traced this to the geometry of (a family of) underlying algebraic curves that encode the experimental data in a neat and compact way. The rigidity of holomorphic modular symmetries follows from the rigidity of the geometry of elliptic curves, and this is what allows us to harvest an infinite number of detailed quantitative predictions.

The complicated special functions needed to describe scaling data arise in a transparent way from the group theory and geometry of these quantum Hall curves, and therefore do not have to be postulated to be the relevant basis for phenomenological “curve fitting”. The RG potential emerges naturally in a geometric setting that complements the phenomenology found in our companion paper².

We have argued that the algebraic geometry of elliptic curves is an efficient way to analyze scaling data, extract the modular symmetries of the transport coefficients, and use this information to fit the given system into the one-dimensional (real) family of curves that may model all universal properties of quantum Hall systems. In particular, we have explained how to find the quantum Hall curve directly from the most robust part of the scaling data. Given such a curve, we have explained how the experimentally relevant phase and flow diagrams are obtained from the geometry of the modular curve that always accompanies an elliptic curve.

Finally, we have discussed the geometry of quantum Hall curves at all RG fixed points. This geometry is singular at the cusps (UV and IR fixed points), and we have described the non-singular geometry at quantum critical points at length. Unfortunately the conformal quantum critical model expected to appear at these points still eludes us.

Acknowledgement

CAL is grateful to Jesus College, Oxford for support, and especially to Andrew Dancer for patient tutorials on algebraic curves and their degenerations.

APPENDIX

We collect here some facts about modular groups and their representations that are used in the main body of this paper.

A. Modular symmetries, functions and forms

The modular group $\Gamma(1) := \text{PSL}(2, \mathbb{Z})$ is generated by $S(\tau) = -1/\tau$ and $T(\tau) = \tau + 1$. They satisfy the algebraic constraints $S^2 = I$ and $(ST)^3 = (TS)^3 = I$, which is the abstract (representation independent) definition of this group.

A *modular function* of weight k for $\Gamma \subset \Gamma(1)$ is a meromorphic function of $\tau \in \mathbb{H}_\infty := \mathbb{H} \cup \{i\infty\}$ with the transformation law

$$f(\gamma(\tau)) = (c\tau + d)^k f(\tau), \quad \gamma(\tau) = \frac{a\tau + b}{c\tau + d},$$

under a modular transformation

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \subsetneq \text{SL}(2, \mathbb{R}) = \text{Isom}(\mathbb{H}).$$

We will denote $f_k|_\gamma(\tau) := (c\tau + d)^{-k} f(\gamma \cdot \tau)$.

Since a translation T^N , for some N , is part of every submodular group, every function will have a Fourier expansion in power of $q_N = e^{2\pi i \tau / N}$. q_N maps \mathbb{H} to the punctured open unit disc and \mathbb{H}_∞ to the open unit disc. For simplicity, we consider here the case $q := q_1$, with Fourier expansion

$$f(\tau) = \sum_{n \in \mathbb{Z}} a_n q^n,$$

with $a_n = 0$ for $n \ll 0$, for a meromorphic function on \mathbb{H}_∞ .

Modular forms are *holomorphic* functions of weight k . The non-positive part $\{a_n\}_{n < 0}$ is called the polar part and is zero for holomorphic forms. In addition, cusp forms are holomorphic forms that vanish at infinity, i.e., $a_0 = 0$. Suitable analytic structure combined with the transformation law typically restrict f_k to be an element in a finite dimensional vector space.

Similarly, for subgroups with several inequivalent cusps, $f_k|_\gamma(\tau)$ is holomorphic in the neighborhood of any inequivalent cusp $\gamma \cdot i\infty$ of \mathbb{F}^{14} .

Clearly if the coefficients $\{a_n\}$ of $f(\tau)$ are real, f satisfies $f(-\bar{\tau}) = \bar{f}(\tau)$. Then $f(\tau)$ is real along the imaginary axis. In fact, then it will also be real on the boundary $\partial \overline{\mathbb{F}}$ associated to Γ , since the boundary of $\overline{\mathbb{F}}$ can be obtained by modular transformations in Γ . The basic j -invariant of $\Gamma(1)$, which classifies elliptic curves, is:

$$\begin{aligned} j(\tau) &= \frac{E_4(\tau)^3}{\Delta(\tau)} \\ &= \frac{1}{q} + 744 + 196684q + 21493769q^2 + \dots, \end{aligned}$$

where

$$\Delta(\tau) = E_4(\tau)^3 - E_6(\tau)^2 \propto \eta^{24}(\tau) \propto \frac{\lambda'^6}{\lambda^4(1-\lambda)^2}$$

is the weight 12 modular discriminant and $E_4(\tau)$ and $E_6(\tau)$ are the weight 4 and 6 Eisenstein functions. Up to a phase the Dedekind η -function

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n)$$

transforms with weight $1/2$, i.e., $\eta(\tau + 1) = e^{i\pi/12} \eta(\tau)$ and $\eta(-1/\tau) = \sqrt{-i\tau} \eta(\tau)$.

The j -function evaluates to every complex number exactly once on the interior of the fundamental domain. This follows from the fact that $j(\tau)$ has a simple pole at $\tau = i\infty$, when τ is in the compactified fundamental domain $\mathcal{M}(\Gamma(1)) \cong \mathbb{CP}^1$.

B. Subgroups at level 2

The modular group has a lot of interesting subgroups that are studied in number theory. Since the quantum Hall plateaux appear at special rationals $\sigma = \sigma_H = p/q \in \mathbb{Q}$, the number-theoretic properties of these groups are of direct relevance. The main congruence subgroup $\Gamma(N)$ at level N is defined as

$$\Gamma(N) = \left\{ \gamma \in \Gamma(1) \mid \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = I \pmod{N} \right\}.$$

The simplest candidate $\Gamma(2) = \langle ST^2S, T^2 \rangle$ has index 6 in $\Gamma(1)$. We are interested in the groups between $\Gamma(1)$ and $\Gamma(2)$.

The coset $\Gamma(1)/\Gamma(2)$ has 6 representatives

$$I \simeq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, S \simeq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, T \simeq \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \quad (30)$$

$$P := ST \simeq \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, TS \simeq \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad (31)$$

$$W := STS \simeq \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}. \quad (32)$$

In fact, since $(ST)^3 = (TS)^3 = S^2 = I$ and $T^2 = I \pmod{\Gamma(2)}$, the coset $\Gamma(1)/\Gamma(2) \cong S_3$, the symmetric group of three objects. Furthermore, since $P^2 = (ST)^2 = T^{-1}S = TS \pmod{\Gamma(2)}$, there are four groups one can define between $\Gamma(1)$ and $\Gamma(2)$, so that $\Gamma(2) \subset \Gamma$. There is a subgroup of index two (sometimes called Γ_P):

$$\Gamma_2 = \{\gamma \in \Gamma(1) \mid \gamma = I, ST, TS \pmod{2}\} = \langle ST, TS \rangle.$$

The other subgroups are of index three and all conjugate to each other

$$\Gamma_0(2) = \{\gamma \in \Gamma(1) \mid \gamma = I, T \pmod{2}\} = \langle ST^2S, T \rangle$$

$$\Gamma^0(2) = \{\gamma \in \Gamma(1) \mid \gamma = I, STS \pmod{2}\} = \langle STS, T^2 \rangle$$

$$\Gamma_\theta(2) = \{\gamma \in \Gamma(1) \mid \gamma = I, S \pmod{2}\} = \langle S, T^2 \rangle.$$

The groups are conjugate via $\Gamma_0(2) = ST^0(2)S$ and $\Gamma^0(2) = T^{-1}\Gamma_\theta(2)T$.

The way the various subgroups above $\Gamma(2)$ treat the rationals can be deduced from how they act on the cusps at 0, 1 and $i\infty$ on their fundamental domain. By labelling $0 \sim e/o$, $1 \sim o/o$ and $i\infty \sim o/e$, these will correspond to the parities of the fractions $p/q \in \mathbb{Q}$ that are grouped into equivalence classes of the cusps $\{0\}$, $\{1\}$ and $\{i\infty\}$ under the group. More concretely, using the transformations of the elements of the group mod 2, we can see what parities of p and q are equivalent modulo the group. We see that the parities of the matrix entries are constrained as follows (“o” is the set of odd integers,

and “e” is the set of even integers):

$$\Gamma_0(2) : \quad c \in e; \quad a = d \in o$$

$$\Gamma^0(2) : \quad b \in e; \quad a = d \in o$$

$$\Gamma_\theta(2) : \quad (a, d \in e; \quad b, c \in o) \vee (a, d \in o; \quad b, c \in e)$$

$$\Gamma(2) : \quad a, d \in o; \quad b, c \in e.$$

The modular group $\Gamma(1)$ does not distinguish between the parities of the fractions $p/q \in \mathbb{Q}$ (all the cusps $\{0, 1, i\infty\}$ are equivalent), whereas the congruence group $\Gamma(2)$ respects the parities of p and q (all the cusps $\{0, 1, i\infty\}$ are inequivalent). The groups of index 2 partition $\{0, 1, i\infty\}$ into two disjoint subsets. In summary, one has the following disjoint sets of cusps in $\partial\mathbb{H} = \mathbb{Q} \cup \{i\infty\}$, that form equivalence classes under the group:

$$\Gamma_0(2) : \quad \{q \in e\} \cup \{q \in o\}$$

$$\Gamma^0(2) : \quad \{p \in e\} \cup \{p \in o\}$$

$$\Gamma_\theta(2) : \quad \{pq \in e\} \cup \{pq \in o\}$$

$$\Gamma(2) : \quad \{p \in o; q \in o\} \cup \{p \in e; q \in o\} \cup \{p \in o; q \in e\}.$$

Here the two sets of inequivalent cusps can be chosen to represent either $\{\oplus\}$ and $\{\ominus\}$. This choice merely inverts the direction of the flow and the corresponding potentials are equivalent up to a sign.

For the index 2 subgroups, these equivalence classes $\{\oplus\}$ and $\{\ominus\}$ of cusps on $\partial\mathbb{H}$ are the set of attractive and repulsive RG fixed points. In addition, there are critical \mathbb{E}_2 saddle points $\{\otimes\}$ on \mathbb{H} , that are critical points of the plateau-plateau transition (see Fig. 1 (d)).

For the group $\Gamma(2)$, there are no \mathbb{E}_2 points, and the critical points $\{\otimes\}$ “condense” to the real axis and form “infinite monkey saddles” \otimes^∞ at 0 or 1 modulo their modular images. Then the set of cusps (and the corresponding rationals) $\{0, 1, i\infty\} \sim \{e/o, o/o, o/e\}$ correspond to the set of fixed points $\{\oplus, \ominus, \otimes^\infty\}$.

Each holomorphic symmetry $\Gamma \subset \Gamma(1)$ determines a unique pair of phase diagrams, depending on whether the fixed point $\infty = 1/0$ is attractive or repulsive. Given this one bit of information Γ completely determines the structure of the phase diagram, including the exact location of all quantum critical points (\otimes), as follows.

If ∞ is an attractive (repulsive) fixed point, then a transition

$$f = p/q = \oplus \leftarrow \otimes \rightarrow \oplus' = p'/q' = f' > f$$

exists for:

$$\Gamma_0(2) \text{ iff } q \text{ and } q' \text{ are even (odd) and } \delta = 2(1)$$

$$\Gamma^0(2) \text{ iff } p \text{ and } p' \text{ are odd (even) and } \delta = 1(2)$$

$$\Gamma_\theta(2) \text{ iff } pq \text{ and } p'q' \text{ are even (odd) and } \delta = 1(2),$$

$$\text{where } \delta = \det(f', f) = p'q - pq'.$$

The subgroup $\Gamma_P = \Gamma_2$ does not distinguish between the parity of the rationals and has \mathbb{E}_3 fixed points at ω and $-\omega^2$ that physically correspond to trifurcated saddle points. Moreover, Γ_2 has no holomorphic modular forms of weight 2.

C. Sub-modular functions and forms

An invariant for the main congruence group $\Gamma(2)$ is

$$\lambda(\tau) = \frac{\theta_2(\tau)^4}{\theta_3(\tau)^4},$$

where the Jacobi theta functions are defined below. The following $q_2 = q^{1/2}$ expansions¹⁸ are useful:

$$\begin{aligned}\lambda(\tau) &= 16q^{1/2} - 128q + 704q^{3/2} - 3072q^2 + \dots \\ \frac{16}{\lambda(\tau)} &= \frac{1}{q^{1/2}} + 8 + 20q^{1/2} - 62q^{3/2} + 216q^{5/2} + \dots,\end{aligned}$$

and $\lambda(0) = 1$, $\lambda(i\infty) = 0$, $\lambda(1) = \infty$.

Invariant functions for the level 2 congruence subgroups $\Gamma_0(2)$, $\Gamma^0(2)$ and $\Gamma_\theta(2)$ with index 2 are, respectively:

$$\begin{aligned}f_1 &= f_T(\tau) = \frac{\Delta(2\tau)}{\Delta(\tau)} \propto \lambda_T = \lambda^{-2}(\lambda - 1) \\ f_2 &= f_R(\tau) = \frac{\Delta(\tau/2)}{\Delta(\tau)} \propto \lambda_R = \lambda(1 - \lambda) \\ f_3 &= f_S(\tau) = \frac{\Delta(\tau)^2}{\Delta(2\tau)\Delta(\tau/2)} \propto \lambda_S = -\lambda(1 - \lambda)^{-2}.\end{aligned}$$

The index 3 subgroup Γ_2 has the invariant function

$$g(\tau) = \left(\frac{\lambda + \omega}{\lambda + \omega^2} \right)^3,$$

but this group is not relevant for the phenomenology of QHE.

The Jacobi theta functions:

$$\begin{aligned}\theta_1(\tau) &= 2 \sum_{n=0}^{\infty} (-1)^n q^{\frac{1}{2}(n+1/2)^2} \\ \theta_2(\tau) &= \sum_{n=-\infty}^{\infty} q^{\frac{1}{2}(n+1/2)^2} \\ \theta_3(\tau) &= \sum_{n=-\infty}^{\infty} q^{\frac{1}{2}n^2} \\ \theta_4(\tau) &= \sum_{n=-\infty}^{\infty} (-1)^n q^{\frac{1}{2}n^2}\end{aligned}$$

are weight 1/2 forms that have no zeros in \mathbb{H} .¹⁵ Clearly $\theta_i(0) = \infty$ or 1, and $\theta_i(i\infty) = 0$ or 1. The Jacobi identity is

$$\theta_3^4 = \theta_4^4 + \theta_2^4.$$

Using Poisson resummation, one can obtain the following

transformation properties

$$\begin{aligned}\theta_2(T(\tau)) &= e^{i\pi/4} \theta_2(\tau) \\ \theta_3(T(\tau)) &= \theta_4(\tau) \\ \theta_4(T(\tau)) &= \theta_3(\tau) \\ \theta_3(S(\tau)) &= \sqrt{-i\tau} \theta_3(\tau) \\ \theta_2(S(\tau)) &= \sqrt{-i\tau} \theta_4(\tau) \\ \theta_4(S(\tau)) &= \sqrt{-i\tau} \theta_2(\tau).\end{aligned}$$

The so-called theta doubling identities are

$$\begin{aligned}2\theta_2(2\tau) \theta_3(2\tau) &= \theta_2^2(\tau) \\ 2\theta_2(2\tau) &= \theta_3^2(\tau) - \theta_4^2(\tau) \\ 2\theta_3(2\tau) &= \theta_3^2(\tau) - \theta_4^2(\tau) \\ \theta_4^2(2\tau) &= \theta_3(\tau) \theta_4(\tau).\end{aligned}$$

The modular λ transforms under T and S as¹⁵

$$\lambda|_T = \frac{\lambda}{\lambda - 1}, \quad \lambda|_S = 1 - \lambda, \quad \lambda'|_S = -\lambda'.$$

The doubling formulas and the relation $\Delta(\tau) = (\theta_2(\tau)\theta_3(\tau)\theta_4(\tau)/2)^8$ imply that:

$$\frac{\Delta(2\tau)}{\Delta(\tau)} = \frac{\theta_2^8(\tau)}{2^8 \theta_3^4(\tau) \theta_4^4(\tau)} = \frac{\lambda^2(\tau)}{2^8 (1 - \lambda(\tau))} = -\frac{1}{2^8 \lambda_T(\tau)}.$$

Other useful relations¹⁵ are:

$$\begin{aligned}\Delta(2\tau) &= \eta^{24}(2\tau) = \theta_2^{16}(\tau) \theta_3^4(\tau) \theta_4^4(\tau) / 2^{16} \\ \lambda_T &= -\frac{\theta_2^8}{\theta_3^4 \theta_4^4}, \quad \lambda_S = \frac{\theta_2^4 \theta_4^4}{\theta_3^8}, \quad \lambda_W = \frac{\theta_2^4 \theta_3^4}{\theta_4^8} \\ f_W(\tau) &= \frac{\Delta(\tau)}{\Delta(\tau/2)} = \frac{1}{2^4} \frac{\theta_2^4 \theta_3^4}{\theta_4^8} \\ f_T^{-1}(\tau) &= 16f_W(2\tau) = 16f_S(2\tau + 1) \\ f_S(\tau) &= f_W(\tau + 1) = f_S(\tau + 2) \\ f_T(-1/\tau) &= -f_W^{-1}(\tau).\end{aligned}$$

The derivative of the modular λ is

$$\lambda'(\tau) = \frac{4i}{\pi} K(k)^2 (1 - \lambda) \lambda,$$

where $K(k)$ is the complete elliptic integral of the first kind

$$K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} = \frac{\pi}{2} \sum_{n=0}^{\infty} \left[\frac{(2n)!}{2^{2n} n!^2} \right]^2 k^{2n}.$$

Some special values are $K(0) = \pi/2$, $K(1) = \infty$, $K(i\infty) = K(\infty) = 0$.

The canonical variable $k(\tau) = \lambda(\tau)^{1/2}$ is called the elliptic modulus. The theta functions are related to elliptic integrals by

$$\begin{aligned}\theta_2(\tau) &= \sqrt{\frac{2kK(k)}{\pi}}, \quad \theta_3(\tau) = \sqrt{\frac{2K(k)}{\pi}} \\ \theta_4(\tau) &= \sqrt{\frac{2k'K(k)}{\pi}}\end{aligned}$$

where $k'^2 = 1 - k^2$ is the complementary modulus. Eq. (7.2.17) in Ref. 15 gives

$$\frac{\lambda'}{1-\lambda} = i\pi\theta_2^4, \quad \frac{\lambda'}{\lambda(1-\lambda)} = i\pi\theta_3^4, \quad \frac{\lambda'}{\lambda} = i\pi\theta_4^4.$$

In particular $K(k) = (\pi/2)\theta_3^2(\tau)$. If we define $K'(k) := K(k')$, then $\exp(\pi i\tau) = \sqrt{q} = \exp(-\pi K'/K)$ and

$$\tau = \frac{iK'(k)}{K(k)} \pmod{\Gamma(2)}.$$

The $\Gamma(2)$ -invariant function λ is essentially the inverse of the elliptic nome $q(m)$ in terms of the parameter $m(q) = \lambda$, and in general elliptic integrals are the inverse functions from the elliptic curve to \mathbb{CP}^1 (see Fig. 2).

-
- ¹ C.A. Lütken, G.G. Ross, Phys. Rev. B 45 (1992) 11837; Phys. Rev. B 48 (1993) 2500; Phys. Lett. A 356 (2006) 382; Phys. Lett. B 653 (2007) 363; Phys. Lett. A 374 (2010) 4700; Nucl. Phys. B 850 (2011) 321.
- ² J. Nissinen, C.A. Lütken, Phys. Rev. B 85 (2012) 155123.
- ³ A.M.M. Pruisken, Nucl. Phys. B 235 (1984) 277; H. Levine, S.B. Libby, A.M.M. Pruisken, Nucl. Phys. B 240 (1984) 30.
- ⁴ A. Altland, B. Simons, *Condensed Matter Field Theory*, Cambridge University Press (2006).
- ⁵ S. Kivelson, D.-H. Lee, S.-C. Zhang, Phys. Rev. B 46 (1992) 2223.
- ⁶ C.P. Burgess, B.P. Dolan, Phys. Rev. B 63 (2001) 155309; Phys. Rev. B 65 (2002) 155323.
- ⁷ R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London (1982).
- ⁸ J.-M. Maillard, S. Boukraa, Ann. Fond. Louis de Broglie 26 (2001) 287.
- ⁹ N. Seiberg, E. Witten, Nucl. Phys. B 426 (1994) 19; Nucl. Phys. B 431 (1994) 484.
- ¹⁰ F. Klein, Math. Annalen 43 (1893) 63.
- ¹¹ Wanli Li, C.L. Vicente, J.S. Xia, W. Pan, D.C. Tsui, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 102 (2009) 216811.
- ¹² D. Husemöller *Elliptic Curves*, Springer, 2nd ed. (2004).
- ¹³ J.H. Silverman, *Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics 106, Springer (2009).
- ¹⁴ F. Diamond, J. Shurman, *A First Course in Modular Forms*, Graduate Texts in Mathematics 228, Springer (2006).
- ¹⁵ R.A. Rankin, *Modular Forms and Functions*, Cambridge University Press (1979).
- ¹⁶ C.H. Clemens, *A Scrapbook of Complex Curve Theory*, Plenum Press, New York (1980).
- ¹⁷ C. P. Burgess, C.A. Lütken, Nucl. Phys. B 500 (1997) 367; Phys. Lett. B 451 (1999) 365; C.A. Lütken, Nucl. Phys. B 759 (2006) 343.
- ¹⁸ M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55, 10th edition (1972).